Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 21 (2019), Issue 4   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Corrigenda and Addenda

Review

Augmented Reality in Emergency Medicine: A Scoping Review

Brendan William Munzer[1], MD; Mohammad Mairaj Khan[1], MD; Barbara Shipman[2], MS; Prashant Mahajan[1], MD, MPH, MBA

[1]Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, United States
[2]Medical School, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Brendan William Munzer, MD
Department of Emergency Medicine
University of Michigan
1501 E Medical Center Dr
Ann Arbor, MI, 48108
United States
Phone: 1 602 403 3296
Email: bmunzer@med.umich.edu

Abstract

Background: Augmented reality is increasingly being investigated for its applications to medical specialties as well as in medical training. Currently, there is little information about its applicability to training and care delivery in the context of emergency medicine.

Objective: The objective of this article is to review current literature related to augmented reality applicable to emergency medicine and its training.

Methods: Through a scoping review utilizing Scopus, MEDLINE, and Embase databases for article searches, we identified articles involving augmented reality that directly involved emergency medicine or was in an area of education or clinical care that could be potentially applied to emergency medicine.

Results: A total of 24 articles were reviewed in detail and were categorized into three groups: user-environment interface, telemedicine and prehospital care, and education and training.

Conclusions: Through analysis of the current literature across fields, we were able to demonstrate that augmented reality has utility and feasibility in clinical care delivery in patient care settings, in operating rooms and inpatient settings, and in education and training of emergency care providers. Additionally, we found that the use of augmented reality for care delivery over distances is feasible, suggesting a role in telehealth. Our results from the review of the literature in emergency medicine and other specialties reveal that further research into the uses of augmented reality will have a substantial role in changing how emergency medicine as a specialty will deliver care and provide education and training.

(J Med Internet Res 2019;21(4):e12368) doi:10.2196/12368

KEYWORDS

augmented reality; emergency medicine; education; telemedicine

Introduction

Since its inception as a specialty, emergency medicine has continually adapted and evolved in the way patient care is delivered as well as in training emergency medicine providers. From the adaptation and utilization of point-of-care ultrasound for procedures and clinical decision making to the methods in which critical care is delivered to patients—both in the prehospital setting and within the emergency department—emergency medical care has leveraged technological advances to improve outcomes for patients [1-3]. Similarly, emergency medicine has been an early adopter for a variety of technology-based education tools, such as simulation and free open-access medical education [4,5].

Augmented reality and virtual reality are two exciting, closely related, but fundamentally different emerging technologies. The central difference between the two technologies is that virtual reality is completely immersive: the headsets must, by necessity, block out the external world. However, augmented reality, by design, maintains the user's connections with the real world. Augmented reality synthesizes the virtual and real; like virtual reality, an augmented reality experience typically involves a headset through which you can view a physical reality that has been augmented or supplemented by computer-generated sensory input such as sound, video, graphics, or GPS data [6]. Both augmented reality and virtual reality are expected to generate US $90 billion in revenue by 2022 [7]. Augmented reality in particular has substantial potential to provide powerful, contextual, and situated learning experiences as well as construct new understanding based upon the user's interactions with virtual objects, which bring underlying data to life. With various augmented reality technologies, one has the ability to perform a wide variety of tasks, including displaying and manipulating information within one's field of view, mapping virtual images to real objects, and video-conferencing [8,9].

Although there have been studies suggesting that there is significant potential for augmented reality in health care, including medical education and surgical subspecialties, at this time there has not been a review of the potential uses for augmented reality in the context of emergency medicine [10,11]. The aim of our paper was to review the medical literature as well as literature in other fields to determine the application of augmented reality in emergency care.

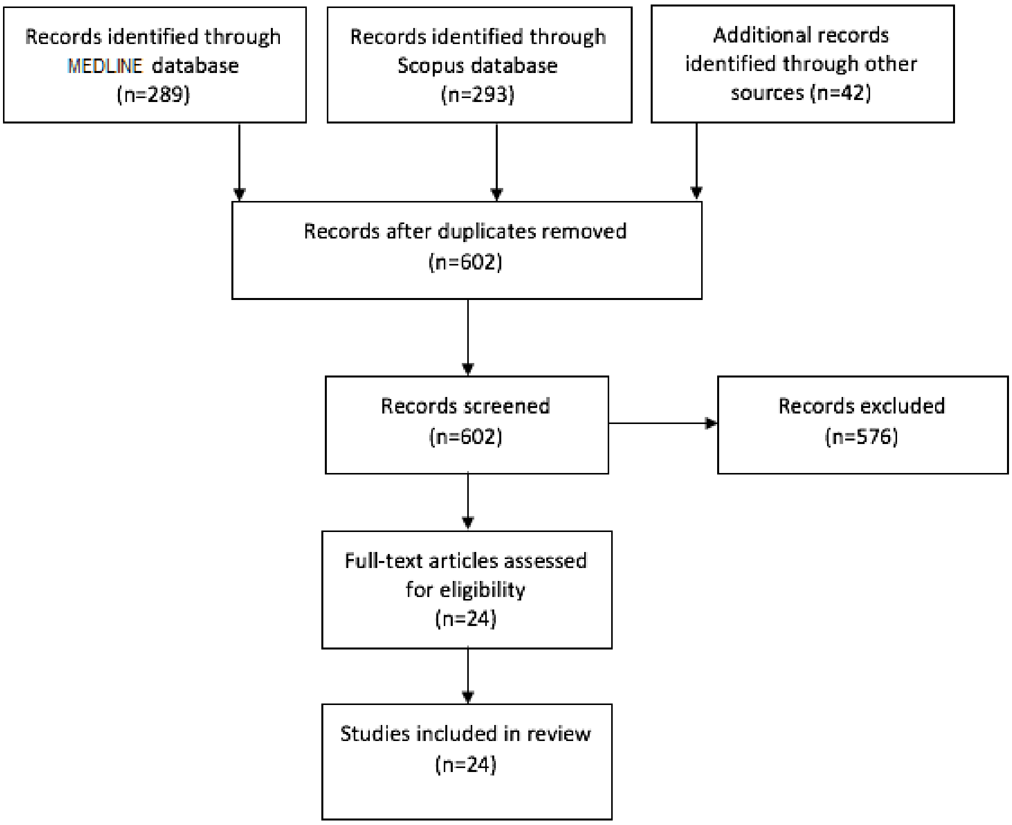
Methods

We conducted a scoping review of both medical and nonmedical journals examining the current use of augmented reality with the assistance of a health sciences informationist (BS).

We elected to perform a scoping review because this area of research is still in its infancy and we therefore sought to quantify the number of articles with potential relevance to emergency medicine. The majority of articles were identified using the Scopus and MEDLINE databases, as well as Embase. We additionally performed a search for relevant articles within the grey literature; however, no records were identified. Following standard Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for study selection, two authors (BWM and MK) evaluated the articles using title and abstract analysis for relevance to emergency medicine and its subspecialties, as detailed in Figure 1 [12]. A total of 20 articles were reviewed by both authors independently to assess interrater reliability, and it was found that there was complete agreement in review. The remaining articles were divided by the two authors evenly and reviewed for relevance. Any articles that were of questionable relevance were reviewed by the two authors independently to come to a consensus. If the authors were split, a third party (PM) was available as an additional reviewer.

Articles were defined as applicable to emergency medicine if there was relevance to a practice within the field of emergency medicine (ie, procedures within the scope of practice of an emergency medicine physician) or its subspecialties (ie, ultrasound or prehospital medicine); if the method in which augmented reality was studied could be applied to emergency medicine (ie, use of an augmented reality headset or novel functionality); and if there was potential educational or teaching value as it relates to emergency medicine. Articles that were excluded from this study were those that were not directly relevant to emergency medicine, were not studying augmented reality, or were not primarily published in the English language.

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



A total of 24 studies were identified as relevant to both augmented reality and emergency medicine. One study was noted to be directly related to emergency medicine. The remaining studies were felt to have topics that were applicable to emergency medicine or its subspecialties. Articles were accessed using the University of Michigan library system.

Results

Overview

A total of 24 articles were selected for full-article review. The main themes within the articles that were found to be relevant to emergency medicine included user-environment interfaces, telemedicine and prehospital care, and education and training. Some articles were found to relate to more than one primary theme. Due to the fact that there were a minimal number of articles directly involving the field of emergency medicine, the majority of the articles reviewed here are from other specialties that could have direct relatability to emergency medicine. Table 1 lists the categories of articles as they relate to emergency medicine [13-36].

User-Environment Interface

Out of 24 articles, 6 (25%) evaluated the interactions between the user and environment. Articles were focused on topics ranging from augmented realism through haptic feedback and anatomic overlaying to usability without interference or distraction and assistance with clinical management.

Siebert et al utilized headsets with the Pediatric Advanced Life Support (PALS) algorithm for defibrillation during a mock code. It was found that there was no increase in time to initial defibrillation and there were fewer errors associated with dosages of defibrillation and medications in those using augmented reality [13]. Usability of the headsets was not discussed in this article. However, in a feasibility trial using Google Glass, Chaballout et al noted concerns regarding distraction through multiple images, overheating of hardware, and difficulties with establishing an Internet connection [14].

Rochlen et al evaluated the feasibility of augmented reality with headsets creating an anatomical layout of an internal jugular vein and carotid artery over a manikin for central venous catheterization insertion [15]. The authors found that, overall, participants found the use of augmented reality easy and enjoyable and found that it did not interfere with the procedural technique.

Abhari et al hypothesized that incorporation of augmented reality into surgical planning would result in improved task performance during resection of neurosurgical tumors. By utilizing a virtual image of the brain overlaying a phantom head, trainees were able to decrease cognitive load in planning a resection and focus more on appropriate technique and surgical maneuvers [16]. Performed in a training environment, this is a first step toward application of augmented reality of procedural tasks in a clinical setting.

Dickey et al utilized a headset with just-in-time training and instruction prior to performing a urologic procedure [17]. Additionally, during the actual procedure, an attending that was not present in the room communicated via headset and video-conferencing and was able to use a cursor to highlight important points within the procedure. Trainees and faculty members in this study felt that the system had educational value, was easy to navigate, and was not distracting. Similarly, Andersen et al describe a system in which orthopedic attendings can directly annotate onto the field of view by using a transparent display on a podium, thereby reducing shifts of focus during a procedure [18].

Table 1. Categorized search results as related to domains of emergency medicine.

Titles of studies in each emergency medicine domain	Journal or conference	First author, year published
User-environment interface		
Adherence to AHA[a] guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: A randomized controlled trial	Journal of Medical Internet Research (JMIR)	Siebert J, 2017 [13]
Feasibility of augmented reality in clinical simulations: Using Google Glass with manikins	JMIR Medical Education	Chaballout B, 2016 [14]
First-person point-of-view augmented reality for central line insertion training: A usability and feasibility study	Simulation in Healthcare	Rochlen L[b], 2017 [15]
Training for planning tumour resection: Augmented reality and human factors	IEEE[c] Transactions on Biomedical Engineering	Abhari K, 2015 [16]
Augmented reality assisted surgery: A urologic training tool	Asian Journal of Andrology	Dickey R[d], 2016 [17]
Avoiding focus shifts in surgical telementoring using an augmented reality transparent display	Studies in Health Technology and Informatics	Andersen D, 2016 [18]
Telemedicine and prehospital care		
Telemedicine supported by augmented reality: An interactive guide for untrained people in performing an ECG[e] test	Biomedical Engineering Online	Bifulco P, 2014 [19]
Augmented reality as a telemedicine platform for remote procedural training	Sensors	Wang S, 2017 [20]
Virtual interactive presence in global surgical education: International collaboration through augmented reality	World Neurosurgery	Davis MC, 2016 [21]
Telementoring: Use of augmented reality in orthopaedic education: AAOS[f] exhibit selection	The Journal of Bone & Joint Surgery	Ponce BA, 2014 [22]
Seven years of clinical experience with teleconsultation in craniomaxillofacial surgery	Journal of Oral and Maxillofacial Surgery	Ewers R, 2005 [23]
Augmented reality assisted surgery: A urologic training tool	Asian Journal of Andrology	Dickey R[d], 2016 [16]
Disaster medicine through Google Glass	European Journal of Emergency Medicine	Carenzo L, 2015 [24]
Technical support by Smart Glasses during a mass casualty incident: A randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine	Journal of Medical Internet Research (JMIR)	Follmann A, 2019 [25]
Education and training		
Design of mobile augmented reality in health care education: A theory-driven framework	JMIR Medical Education	Zhu E, 2015 [26]
Systematic review on the effectiveness of augmented reality applications in medical training	Surgical Endoscopy	Barsom EZ, 2016 [27]
Personalized augmented reality for anatomy education	Clinical Anatomy	Ma M, 2016 [28]
Augmented reality for anatomical education	Journal of Visual Communication in Medicine	Thomas RG, 2010 [29]
The effectiveness of virtual and augmented reality in health sciences and medical anatomy	Anatomical Sciences Education	Moro C, 2017 [30]
Modification of commercial force feedback hardware for needle insertion simulation	Studies in Health Technology and Informatics	Coles TR, 2011 [31]
Augmented reality for teaching endotracheal intubation: MR[g] imaging to create anatomically correct models	Annual AMIA[h] Symposium Proceedings	Kerner KF, 2003 [32]
First-person point-of-view augmented reality for central line insertion training: A usability and feasibility study	Simulation in Healthcare	Rochlen L[b], 2017 [15]
Advanced training methods using an augmented reality ultrasound simulator	IEEE 2009 International Symposium on Mixed and Augmented Reality	Blum T, 2009 [33]
An augmented reality simulator for ultrasound guided needle placement training	Medical & Biological Engineering & Computing	Magee D, 2007 [34]

Titles of studies in each emergency medicine domain	Journal or conference	First author, year published
Piloting augmented reality technology to enhance realism in clinical simulation	Computers, Informatics, Nursing	Vaughn J, 2016 [35]
An investigation of university students' collaborative inquiry learning behaviors in an augmented reality simulation and a traditional simulation	Journal of Science Education and Technology	Wang H-Y, 2014 [36]

[a] AHA: American Heart Association.

[b] Cross-listed in User-environment interface and Education and training.

[c] IEEE: Institute of Electrical and Electronics Engineers.

[d] Cross-listed in User-environment interface and Telemedicine and prehospital.

[e] ECG: electrocardiogram.

[f] AAOS: American Academy of Orthopaedic Surgeons.

[g] MR: magnetic resonance.

[h] AMIA: American Medical Informatics Association.

Telemedicine and Prehospital Care

There were 8 out of 24 articles (33%) focusing on the utility of augmented reality in the prehospital environment or through remote locations. Articles in this section focused on feasibility of providing instruction from a remote location, accuracy, and timeliness of feedback provided. Additionally, this section looks at articles that utilize augmented reality for prehospital care and triage in disaster management.

Augmented reality-supported telemedicine was used by both Bifulco et al and Wang et al to train novices in procedures [19,20]. Bifulco et al demonstrated that it was feasible to use augmented reality to assist with training novices to correctly perform an electrocardiogram (ECG) with minimal errors through the use of cameras, headsets, and annotations. Similarly, Wang et al used the Microsoft HoloLens to remotely instruct novices on point-of-care ultrasound. A facilitator guided learners through remote image capture of hand positions, providing instruction on complex maneuvers. Overall, results from this study demonstrated feasibility in an augmented reality platform for telemedicine as well as feasibility of the HoloLens as a headset.

Davis et al used an iPad-based virtual and augmented reality-based tool to communicate from Birmingham, Alabama, to Ho Chi Minh City, Vietnam, and provide both visual and verbal cues to assist with neurosurgical procedures [21]. The average length of delay in communication was 237 milliseconds and there was no disruption in the quality of the image transmitted. Similarly, Ponce et al used the same tool to demonstrate effectiveness with orthopedic residents performing arthroscopic shoulder surgery [22]. In this study, both residents and attendings felt that the use of augmented reality was educational, was easy to use, and provided immediate feedback to the trainee without causing a significant lag or interference with the procedure. Ewers et al assisted in 50 procedures over a 7-year period using augmented reality, noting that 48 of 60 (80%) transmission attempts were successful with good image quality [23]. As mentioned above, Dickey et al used remote location attendings to provide assistance and guidance for residents performing a urologic procedure [17].

Carenzo et al demonstrated an application for augmented reality in disaster triage using Google Glass. Through the development of an app based upon geomarked quick-response (QR) codes, electronic triaging of disaster patients was performed in a mock triage event. Using hands-free technology and Google Glass as a QR scanner, emergency medical services providers were directed to appropriate patients in an efficient manner [24]. Follmann et al compared traditional triaging during a simulated mass casualty event to the use of Smart Glasses through both an interactive triage algorithm app and telemedical support. The study found that while traditional triaging was faster, triaging using both an interactive Smart Glasses app and a physician with triage experience available by telemedicine was significantly more accurate [25].

Education and Training

Out of 24 articles, 12 (50%) regarding augmented reality with relation to emergency medicine had a focus on education and training. Several of the articles were systematic reviews of the use of augmented reality in medical education as well as conceptual frameworks that can be applied to multiple specialties. Education articles focused on both procedural training as well as clinical decision making in a simulated environment.

Zhu et al proposed a framework when developing augmented reality applications, based upon three different learning theories: situated learning, experiential learning, and transformative learning. By incorporating augmented reality into a realistic learning environment that is insulated from the high-stakes patient care location, the authors propose that the learning that takes place in an augmented reality-assisted environment can inform real practice, while allowing for reflection through experiences [26]. In considering this framework for development of augmented reality applications, there is an emphasis on developing multiple learning modalities within an environment.

Barsom et al performed a systematic review of augmented reality systems that exist that have been validated within a training environment. With specific relation to emergency medicine, two separate echocardiography augmented reality applications were reviewed: one that was felt to be highly realistic, suggesting face validity, and the other determined to have construct validity as determined by different levels of competency for experts, intermediate learners, and novices [27].

Several author groups have looked at the use of augmented reality in anatomical education. Ma et al utilized computed tomography (CT)-developed imaging overlaid over learners for educational benefit. Learners then stood in front of a front-facing video camera, overlaying anatomical images on their bodies [28]. Users found this to be an enjoyable way to relate anatomy to a real situation. Thomas et al used an augmented reality model for developing 3D brains for anatomical study, while also adding a tactile component for interactive learning [29]. Their model allowed for annotation and manipulation in a 3D environment without relying on conventional dissection. Moro et al compared virtual reality, augmented reality, and tablet-based methods for teaching skull anatomy [30]. The study found that virtual and augmented reality were equally effective in teaching anatomy but promoted increased engagement and learner immersion due to the tactile and 3D nature of these learning modes. Virtual reality-based learning was associated with increased episodes of dizziness compared to the other two modalities.

One of the most common utilities for augmented reality currently being investigated is procedural learning. Coles et al used a simplified hydraulic pump to simulate a femoral pulse for needle placement in an interventional radiology. Using augmented reality, the authors simulated the clinical environment as well as the actual needle insertion [31]. The study found that learners felt the model was realistic and was educational. Kerner et al used magnetic resonance-based images to create anatomically correct airways to teach learners appropriate technique for endotracheal intubation [32]. By overlaying images using augmented reality, facilitators were able to demonstrate the changes in airway patency with changes in positioning, such as the sniffing position, extension, and hyperflexion. Rochlen et al used an anatomical overlay to teach learners appropriate technique for central venous catheterization [15].

In addition to the above procedural skills, augmented reality has been used to simulate ultrasound training as well as ultrasound-guided procedures. Blum et al used CT images with a phantom ultrasound simulator to create ultrasound images from CT slices [33]. Sensors placed on the probe, as well as the phantom manikin, interact to simulate ultrasound images from the existing CT scan. Magee et al used augmented reality to simulate ultrasound images and anatomy for interventional radiology procedures [34]. Participants felt that there was face validity in the simulated procedure, in that the augmented reality needle placement accurately replicated the look of the true procedure. It is noted in this article that there was not a mechanism for tactile feedback and the procedure was therefore not considered to feel realistic by participants.

Vaughn et al piloted augmented reality technology using Google Glass with nursing students to add additional realism to a simulated manikin scenario [35]. Learners felt that a simulated patient video provided an added level of realism to the scenario, giving them confidence when they successfully completed the scenario. Wang et al compared students undergoing traditional simulation scenarios with students participating in augmented reality simulation scenarios. Students manipulated virtual 3D cubes to teach them about physics concepts [36]. The behaviors and interactions of the groups were compared, and it was found that students participating in augmented reality simulation scenarios were more supportive of each other in a collaborative environment, though the sample size was small and significance was not reached.

Discussion

Principal Findings

Our scoping review revealed that a majority of studies that have been conducted can be categorized into one of three distinct themes: *user-environment interfaces*, *telemedicine and prehospital care*, and *medical education and training*. Furthermore, our analysis revealed that most of the studies were observational in nature and were conducted on small cohorts of participants. Despite these limitations, we believe that augmented reality has a substantial potential to impact emergency care but would need to be investigated in a more robust manner.

The studies cited above have shown that augmented reality can provide useful user-environmental interfaces in a health care context. Emergency department care demands real-time decision making that requires analysis of objective and subjective data that currently is displayed on computer screens and monitors that divert provider attention away from patients. Live vital sign information along with laboratory data and imaging can all be displayed on an augmented reality platform that limits provider distraction from his or her patient. Furthermore, in the same way that augmented reality has been used in surgical settings in the context of procedures and procedural planning, augmented reality can be used during live procedures in the emergency department. For example, ultrasound imaging can be added to a visual interface that limits providers from looking away from their patient and their needle and devices [15,33]. Expert guidance during procedures and telemedicine can also be delivered via such a platform [17,18,25].

Future Directions

Augmented reality has potential in both the prehospital and telemedicine settings. Similar to the example above, prehospital providers can benefit from expert guidance with visual instructions relayed to their augmented reality devices. Such devices can also provide visual and audio streams for emergency departments who are expecting the arrival of critical patients to help guide their initial resuscitation. Augmented reality can provide an avenue through which rural and international emergency departments can be given guidance on specialized care and procedures. Similarly, augmented reality is a modality that has the potential to revolutionize medical training and education. Augmented reality grants the potential to create hyper-realistic simulations that can give learners a safe environment in which to learn various procedures and skills.

As an emerging technology, augmented reality has substantial potential to enhance quality of care, improve access, and reduce errors especially when contextualized to acute care. To date, there are no studies that evaluate the effect of augmented reality on patient safety in real time. It will be important in developing this technology to evaluate not only its impact on skill acquisition and telemedicine, but to evaluate the use of augmented reality on patient safety and clinical outcomes.

One of the major advantages of augmented reality is that, unlike other computer interfaces, it analyzes and displays information within an individual's field of view, thus allowing information to be utilized in real time. In emergency medicine, appropriate delivery of care often requires quick and easy access to pertinent data that is currently delivered via monitors and computer screens. This can be distracting and lead to inefficiencies which augmented reality can minimize. Additionally, during surgery and other procedures, augmented reality can provide a modality through which information can be delivered to providers in a novel way. Medical education can also benefit from augmented reality, as simulation and basic anatomic and physiological concepts can be displayed in ways that 2D screens cannot achieve.

Prior to augmented reality, the primary modality of interfacing with such information has been through different types of screens and monitors. Augmented reality requires considerable computing and hardware capabilities, and therefore wide-scale development of augmented reality has been limited until recently. Augmented reality software and hardware are now being actively developed by major companies, such as Apple, Microsoft, Facebook, and Google, and is set to become much more ubiquitous in the near future [37].

Historically, medicine as a field has taken time to adapt emerging general-purpose technologies. Despite the widespread use of digital records in other fields, electronic medical records have only recently become mandated [38].

The success of augmented reality as a data interface for providers in the emergency department will be based upon a variety of factors. Cost, software development, and portability will all be crucial determining factors on whether or not augmented reality will find a place in the emergency department of the future. Devices such as the Microsoft HoloLens provide the functionality necessary for widespread application but may not be realistic to wear in the emergency department environment yet. Currently, companies such as Apple are investigating the use of augmented reality as an extension of their mobile phones [37]. As this technology advances, more studies will be needed to evaluate the different modalities for which augmented reality can be used. Most of these factors are dependent on advancements in the private sector. Future augmented reality devices will likely also have to be able to connect and interface with other medical devices, such as vital monitors, ultrasound machines, and emergency medical record systems, in order to be successful. This will require software standards to be developed. Finally, augmented reality will have to prove beneficial to patient outcomes, provider education, emergency department efficiency, and overall cost reduction to convince health care systems to invest in such technologies.

Limitations

Our review did have limitations that are worth discussing. This review was limited to studies from two databases—MEDLINE and Scopus—and at this time there is no way to evaluate augmented reality applications that are being developed by private enterprise. Our initial aim was to find all studies that could be generalized to emergency medicine as a specialty and did not focus on one specific application within the field of emergency medicine. However, as augmented reality is a nascent technology, research involving it is limited, and a more specific clinical question would not have yielded enough papers to conduct such a review. Another limitation is that the included studies described small-scale implementations of augmented reality and use various different augmented reality platforms in their research. Lastly, the majority of these studies did not use objective data to evaluate augmented reality effectiveness. As augmented reality becomes more ubiquitous, we believe that these limitations will pose less of an issue.

In addition to the limitations described above, there are limitations to the current devices on the market. As this is a new technology, the wearables often have low battery life, less-than-ideal camera quality, and the potential for poor wireless connection. Furthermore, there is little information on usability in a health care setting with regard to activities such as sanitation and usability. Given that augmented reality is an emerging technology, the hand swipes and gestures may be cumbersome for health care providers, akin to when advanced-feature mobile phones were first introduced. As with any new technology, it can take time to develop the requisite technology that will suit the needs of health care. These limitations also stress the importance of quality future research looking at the feasibility of using augmented reality in novel settings such as the emergency department, especially as it relates to the devices and user interactions. Future studies should not only investigate this but compare it to current technology to determine its benefits.

Conclusions

An important conclusion that can be drawn from this review is that any research involving augmented reality can be used to inform other applications across various disciplines and areas. We believe that augmented reality has a plethora of applications in emergency medicine and possesses the potential to revolutionize delivery of patient care and medical education. However, further research is needed and should focus on augmented reality's effectiveness in tackling the more specific problems faced in medical settings. Though our review has focused primarily on the emergency department, many of these challenges may be relatable to other clinical environments in medicine. Our review found that although current research is sparse in augmented reality and its application for clinical care delivery and education, augmented reality has substantial potential to change the paradigm of care for acutely ill and injured patients and populations.

Conflicts of Interest

None declared.

References

1. Moore CL, Copel JA. Point-of-care ultrasonography. N Engl J Med 2011 Feb 24;364(8):749-757. [doi: 10.1056/NEJMra0909487] [Medline: 21345104]
2. Rivers EP, Nguyen HB, Huang DT, Donnino MW. Critical care and emergency medicine. Curr Opin Crit Care 2002 Dec;8(6):600-606. [Medline: 12454549]
3. Seymour CW, Rea TD, Kahn JM, Walkey AJ, Yealy DM, Angus DC. Severe sepsis in pre-hospital emergency care: Analysis of incidence, care, and outcome. Am J Respir Crit Care Med 2012 Dec 15;186(12):1264-1271 [FREE Full text] [doi: 10.1164/rccm.201204-0713OC] [Medline: 23087028]
4. Jones F, Passos-Neto CE, Melro Braghiroli OF. Simulation in medical education: Brief history and methodology. Princ Pract Clin Res 2015;1(2):56-63 ISSN: 2378-1890.
5. Nickson CP, Cadogan MD. Free Open Access Medical education (FOAM) for the emergency physician. Emerg Med Australas 2014 Feb;26(1):76-83. [doi: 10.1111/1742-6723.12191] [Medline: 24495067]
6. Thomas C. Download3K.com. 2015 Jan 02. The difference between augmented reality and virtual reality URL: https://www.download3k.com/articles/The-Difference-Between-Augmented-Reality-and-Virtual-Reality-00992 [accessed 2019-03-12] [WebCite Cache ID 76p5scARK]
7. Digi-Capital. 2018 Jan 26. Ubiquitous $90 billion AR to dominate focused $15 billion VR by 2022 URL: https://www.digi-capital.com/news/2018/01/ubiquitous-90-billion-ar-to-dominate-focused-15-billion-vr-by-2022/ [accessed 2019-03-12] [WebCite Cache ID 76p5wJSit]
8. Berryman DR. Augmented reality: A review. Med Ref Serv Q 2012;31(2):212-218. [doi: 10.1080/02763869.2012.670604] [Medline: 22559183]
9. Brigham TJ. Reality check: Basics of augmented, virtual, and mixed reality. Med Ref Serv Q 2017;36(2):171-178. [doi: 10.1080/02763869.2017.1293987] [Medline: 28453428]
10. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: An integrative review. PeerJ 2014;2:e469 [FREE Full text] [doi: 10.7717/peerj.469] [Medline: 25071992]
11. Kim Y, Kim H, Kim YO. Virtual reality and augmented reality in plastic surgery: A review. Arch Plast Surg 2017 May;44(3):179-187 [FREE Full text] [doi: 10.5999/aps.2017.44.3.179] [Medline: 28573091]
12. 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]
13. Siebert JN, Ehrler F, Gervaix A, Haddad K, Lacroix L, Schrurs P, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: A randomized controlled trial. J Med Internet Res 2017 Dec 29;19(5):e183 [FREE Full text] [doi: 10.2196/jmir.7379] [Medline: 28554878]
14. Chaballout B, Molloy M, Vaughn J, Brisson III R, Shaw R. Feasibility of augmented reality in clinical simulations: Using Google Glass with manikins. JMIR Med Educ 2016 Mar 07;2(1):e2 [FREE Full text] [doi: 10.2196/mededu.5159] [Medline: 27731862]
15. Rochlen LR, Levine R, Tait AR. First-person point-of-view augmented reality for central line insertion training: A usability and feasibility study. Simul Healthc 2017 Feb;12(1):57-62 [FREE Full text] [doi: 10.1097/SIH.0000000000000185] [Medline: 27930431]
16. Abhari K, Baxter JSH, Chen ECS, Khan AR, Peters TM, de Ribaupierre S, et al. Training for planning tumour resection: Augmented reality and human factors. IEEE Trans Biomed Eng 2015 Jun;62(6):1466-1477. [doi: 10.1109/TBME.2014.2385874] [Medline: 25546854]
17. Dickey RM, Srikishen N, Lipshultz LI, Spiess PE, Carrion RE, Hakky TS. Augmented reality assisted surgery: A urologic training tool. Asian J Androl 2016;18(5):732-734 [FREE Full text] [doi: 10.4103/1008-682X.166436] [Medline: 26620455]
18. Andersen D, Popescu V, Cabrera ME, Shanghavi A, Gomez G, Marley S, et al. Avoiding focus shifts in surgical telementoring using an augmented reality transparent display. Stud Health Technol Inform 2016;220:9-14. [Medline: 27046545]
19. Bifulco P, Narducci F, Vertucci R, Ambruosi P, Cesarelli M, Romano M. Telemedicine supported by augmented reality: An interactive guide for untrained people in performing an ECG test. Biomed Eng Online 2014 Nov 21;13:153 [FREE Full text] [doi: 10.1186/1475-925X-13-153] [Medline: 25413448]
20. Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, et al. Augmented reality as a telemedicine platform for remote procedural training. Sensors (Basel) 2017 Oct 10;17(10):2294 [FREE Full text] [doi: 10.3390/s17102294] [Medline: 28994720]
21. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: International collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
22. Ponce BA, Jennings JK, Clay TB, May MB, Huisingh C, Sheppard ED. Telementoring: Use of augmented reality in orthopaedic education: AAOS exhibit selection. J Bone Joint Surg Am 2014 May 21;96(10):e84. [doi: 10.2106/JBJS.M.00928] [Medline: 24875036]

23. Ewers R, Schicho K, Wagner A, Undt G, Seemann R, Figl M, et al. Seven years of clinical experience with teleconsultation in craniomaxillofacial surgery. J Oral Maxillofac Surg 2005 Oct;63(10):1447-1454. [doi: 10.1016/j.joms.2005.06.020] [Medline: 16182912]
24. Carenzo L, Barra FL, Ingrassia PL, Colombo D, Costa A, Della Corte F. Disaster medicine through Google Glass. Eur J Emerg Med 2015 Jun;22(3):222-225. [doi: 10.1097/MEJ.0000000000000229] [Medline: 25460812]
25. Follmann A, Ohligs M, Hochhausen N, Beckers S, Rossaint R, Czaplik M. Technical support by Smart Glasses during a mass casualty incident: A randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. J Med Internet Res 2019 Jan 03;21(1):e11939 [FREE Full text] [doi: 10.2196/11939] [Medline: 30609988]
26. Zhu E, Lilienthal A, Shluzas LA, Masiello I, Zary N. Design of mobile augmented reality in health care education: A theory-driven framework. JMIR Med Educ 2015 Sep 18;1(2):e10 [FREE Full text] [doi: 10.2196/mededu.4443] [Medline: 27731839]
27. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Dec;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
28. Ma M, Fallavollita P, Seelbach I, Von Der Heide AM, Euler E, Waschke J, et al. Personalized augmented reality for anatomy education. Clin Anat 2016 May;29(4):446-453. [doi: 10.1002/ca.22675] [Medline: 26646315]
29. Thomas RG, John NW, Delieu JM. Augmented reality for anatomical education. J Vis Commun Med 2010 Mar;33(1):6-15. [doi: 10.3109/17453050903557359] [Medline: 20297908]
30. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
31. Coles TR, John NW, Sofia G, Gould DA, Caldwell DG. Modification of commercial force feedback hardware for needle insertion simulation. Stud Health Technol Inform 2011;163:135-137. [Medline: 21335776]
32. Kerner KF, Imielinska C, Rolland J, Tang H. Augmented reality for teaching endotracheal intubation: MR imaging to create anatomically correct models. AMIA Annu Symp Proc 2003:888 [FREE Full text] [Medline: 14728393]
33. Blum T, Heining S, Kutter O, Navab N. Advanced training methods using an augmented reality ultrasound simulator. In: Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality. 2009 Presented at: 8th IEEE International Symposium on Mixed and Augmented Reality; October 19-22, 2009; Orlando, FL p. 177-178. [doi: 10.1109/ISMAR.2009.5336476]
34. Magee D, Zhu Y, Ratnalingam R, Gardner P, Kessel D. An augmented reality simulator for ultrasound guided needle placement training. Med Biol Eng Comput 2007 Oct;45(10):957-967. [doi: 10.1007/s11517-007-0231-9] [Medline: 17653784]
35. Vaughn J, Lister M, Shaw RJ. Piloting augmented reality technology to enhance realism in clinical simulation. Comput Inform Nurs 2016 Sep;34(9):402-405. [doi: 10.1097/CIN.0000000000000251] [Medline: 27258807]
36. Wang H, Been-Lirn DH, Li N. An investigation of university students' collaborative inquiry learning behaviors in an augmented reality simulation and a traditional simulation. J Sci Educ Technol 2014;23(23):682-691. [doi: 10.1007/s]
37. Vanian J. Fortune. 2017 Jun 07. This is the next battleground for Apple, Google, Facebook, and Microsoft URL: http://fortune.com/2017/06/07/apple-google-facebook-microsoft-are-battling-augmented-reality/ [accessed 2019-03-12] [WebCite Cache ID 76p9ghEKy]
38. Webster PC. Electronic health records a "strong priority" for US government. CMAJ 2010 May 18;182(8):E315-E316 [FREE Full text] [doi: 10.1503/cmaj.109-3218] [Medline: 20351129]

Abbreviations

AAOS: American Academy of Orthopaedic Surgeons
AHA: American Heart Association
AMIA: American Medical Informatics Association
CT: computed tomography
ECG: electrocardiogram
IEEE: Institute of Electrical and Electronics Engineers
MR: magnetic resonance
PALS: Pediatric Advanced Life Support
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QR: quick response

Edited by C Eichenberg; submitted 30.09.18; peer-reviewed by K Kowalewski, A Follmann; comments to author 29.10.18; revised version received 31.01.19; accepted 28.02.19; published 17.04.19.

Please cite as:
Munzer BW, Khan MM, Shipman B, Mahajan P
Augmented Reality in Emergency Medicine: A Scoping Review
J Med Internet Res 2019;21(4):e12368
URL: http://www.jmir.org/2019/4/e12368/
doi:10.2196/12368
PMID:30994463

Review

The Use of Cancer-Specific Patient-Centered Technologies Among Underserved Populations in the United States: Systematic Review

Will L Tarver[1,2], DrPH, MLIS; David A Haggstrom[1,3,4], MD, MAS

[1]VA Health Services Research and Development, Center for Health Information & Communication, Richard L Roudebush VA Medical Center, Indianapolis, IN, United States

[2]Department of Health Policy & Management, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

[3]Division of General Internal Medicine and Geriatrics, School of Medicine, Indiana University, Indianapolis, IN, United States

[4]Center for Health Services Research, Regenstrief Institute, Indianapolis, IN, United States

Corresponding Author:
Will L Tarver, DrPH, MLIS
VA Health Services Research and Development
Center for Health Information & Communication
Richard L Roudebush VA Medical Center
1481 W 10th St
Indianapolis, IN, 46202
United States
Phone: 1 317 988 2616
Email: wltarver@iu.edu

Abstract

Background: In the United States, more than 1.6 million new cases of cancer are estimated to be diagnosed each year. However, the burden of cancer among the US population is not shared equally, with racial and ethnic minorities and lower-income populations having a higher cancer burden compared with their counterparts. For example, African Americans have the highest mortality rates and shortest survival rates for most cancers compared with other racial or ethnic groups in the United States. A wide range of technologies (eg, internet-based [electronic health, eHealth] technologies, mobile [mobile health, mHealth] apps, and telemedicine) available to patients are designed to improve their access to care and empower them to participate actively in their care, providing a means to reduce health care disparities; however, little is known of their use among underserved populations.

Objective: The aim of this study was to systematically review the current evidence on the use of cancer-specific patient-centered technologies among various underserved populations.

Methods: Computer-based search was conducted in the following academic databases: (1) PubMed (cancer subset), (2) MEDLINE, (3) PsycINFO, and (4) CINAHL. We included studies that were peer-reviewed, published in the English language, and conducted in the United States. Each study was individually assessed for relevance, with any disagreements being reconciled by consensus. We used a 3-step inclusion process in which we examined study titles, abstracts, and full-text papers for assessment of inclusion criteria. We systematically extracted information from each paper meeting our inclusion criteria.

Results: This review includes 71 papers that use patient-centered technologies that primarily targeted African Americans (n=31), rural populations (n=14), and Hispanics (n=12). A majority of studies used eHealth technologies (n=41) finding them to be leading sources of cancer-related health information and significantly improving outcomes such as screening among nonadherent individuals and increasing knowledge about cancer and cancer screening. Studies on mHealth found that participants reported overall favorable responses to receiving health information via short message service (SMS) text message; however, challenges were experienced with respect to lack of knowledge of how to text among some participants. More complex mobile technologies (eg, a tablet-based risk assessment tool) were also found favorable to use and acceptable among underserved populations; however, they also resulted in more significant barriers, for example, participants expressed concerns regarding security and unfamiliarity with the technology and preferred further instruction and assistance in its use.

Conclusions: There is a growing body of literature exploring patient-centered technology and its influence on care of underserved populations. In this review, we find that these technologies seem to be effective, especially when tailored, in improving patient and care-related outcomes. Despite the potential of patient-centered technologies and the receptivity of underserved populations, challenges still exist with respect to their effective use and usability.

(J Med Internet Res 2019;21(4):e10256) doi:10.2196/10256

KEYWORDS

underserved populations; medical informatics; cancer

Introduction

Background

In the United States, more than 1.6 million new cases of cancer are estimated to be diagnosed each year [1]; however, the burden of cancer among the US population is not shared equally. Medically underserved populations are defined as groups with economic, cultural, or linguistic barriers to medical care services [2]. These groups include racial and ethnic minorities and individuals of lower socioeconomic status [3] who have a higher cancer burden compared with their counterparts, which can be partially attributed to differences in the access to, and quality of, care they receive [4-6]. A wide range of technologies is available to patients, which have the potential to improve access to care and empower individuals to participate more actively in their care [7,8]. These technologies include personal health records (PHRs) [9], internet-based (eHealth) technologies [10,11], mobile (mHealth) apps [12], and telemedicine [13]. For example, there is evidence that patient-centered technologies (also commonly referred to as *consumer health information technologies*) provide patient-centered care by increasing patients' quality of health care [14], improving communication with providers [13,15-17], providing tailored education and lifestyle messages [14,18], and promoting self-management of health care [19]. Among other cancer health promotion activities, these technologies educate individuals on the benefits of cancer screening, enable individuals to receive reminders for cancer screening and follow-up, and provide tailored decision aids for cancer care. These health information technologies have also been proposed as a means to reduce health care disparities [8,20-22]. The Institute of Medicine identified the internet and computers as critical vehicles to deliver health information to reach diverse populations of cancer patients and survivors [23,24], including low-literacy and low-income African Americans [25-28]. Research is crucial to understand the use and impact of these technologies among underserved populations for the purposes of cancer health promotion, but the medical literature has not been systematically reviewed to understand these patterns or outcomes.

To date, reviews on the use of patient-centered technologies have largely focused on the general population. For example, a recent review by Kim and Nahm [29] found several benefits to the use of patient-accessible PHRs, including consumer empowerment, improved patient-provider communication, increased access to data during times of emergency, improved chronic disease management, and increased likelihood of behavior change. Several concerns were also raised regarding the broader dissemination of PHRs, including data privacy and security, data accuracy, health literacy, and the digital divide. With regard to mobile technologies, Krishna et al [30] conducted a review and found significant improvements in medication adherence, smoking quit rates, self-efficacy, and other health outcomes (eg, asthma symptoms, blood sugar control, and stress levels). Limited attention has also been given to the potential of patient-accessible PHRs among specific disease classes [31,32]; however, Price et al [33] found PHR interventions targeting asthma, diabetes, fertility, glaucoma, HIV, hyperlipidemia, and hypertension (but not cancer) to have beneficial effects such as better quality of care, improved access to care, and increased productivity. In another review, Bennet et al [34] found that racial or ethnic minority populations have been targeted with interventions to facilitate weight loss. Overall, internet-based technologies (eHealth) were only able to affect short-term weight loss, whereas mobile technologies (mHealth) provided no benefit [34]. Montague et al [35] also found that technologies can positively affect the health of the underserved if they are effectively tailored, but little is known about how to effectively tailor cancer-specific technologies to this population. Although eHealth and mHealth studies have also explored the use and adoption of these technologies [36-40] and have been shown to increase screening for cancer [41] and knowledge of cancer and cancer screening [42], these studies have not been systematically reviewed and synthesized to date.

The purpose of this study is to systematically review current evidence on the use of cancer-specific patient-centered technologies among underserved populations. This review contributes to both the informatics and cancer health disparities literature by seeking to address the following issues: (1) to understand the effect or impact of patient-centered technologies on the health or health care outcomes of underserved populations, (2) to understand the use, usability, and acceptance of patient-centered technologies and efforts to tailor their design to improve cancer care among underserved populations, (3) to understand the barriers and facilitators to patient-centered technology use for different populations, and (4) to propose directions for future research based on the current literature.

Conceptual Framework

Patient-centered technology use by underserved populations is influenced by multiple factors. For purposes of this review, we have adapted an existing health services research framework to organize the factors that influence the use and acceptance of information technology among individuals. Originally developed with the organization in mind, the unified theory of acceptance and use of technology (UTAUT) [43] sought to understand the critical factors related to the prediction of behavioral intention to use technologies within the organizational context.

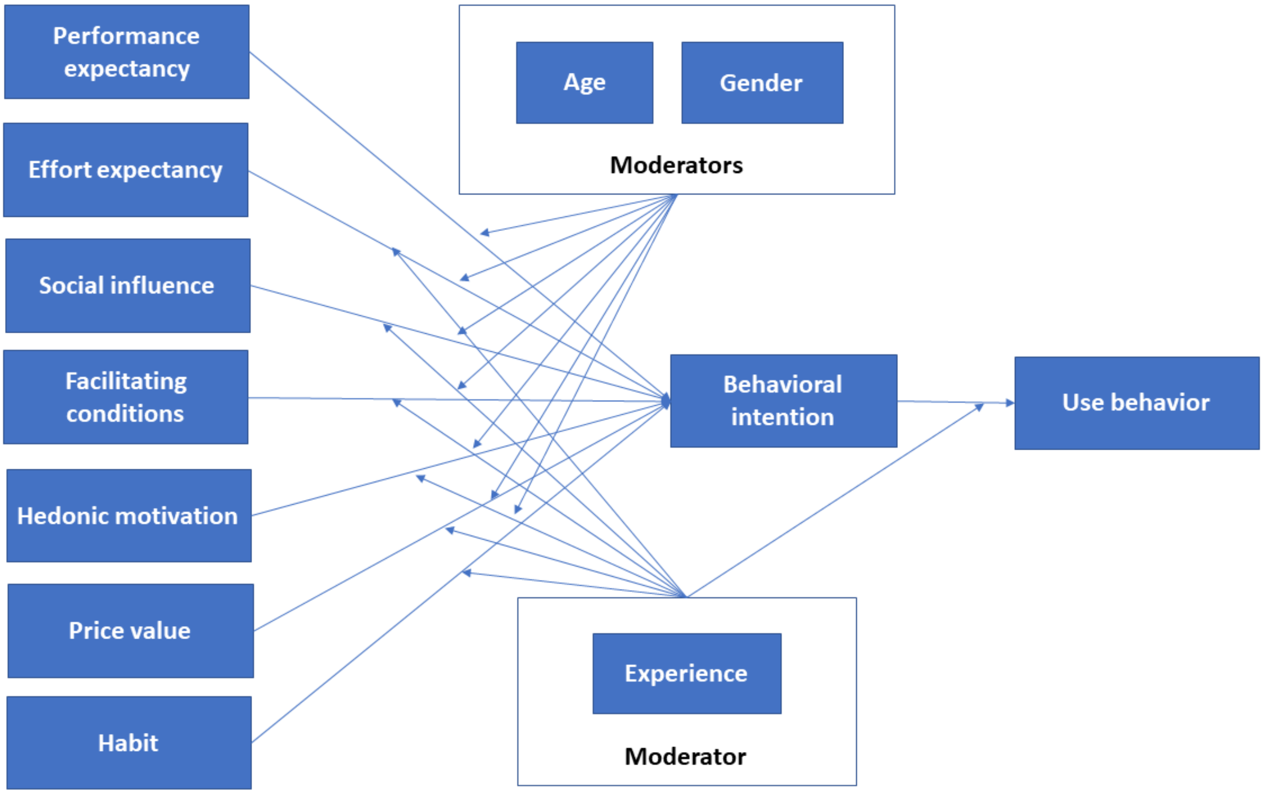
According to the UTAUT, there are 3 constructs (ie, performance expectancy, effort expectancy, and social influence) that are considered direct determinants of the intention to use technology (see Table 1 and Figure 1). *Performance expectancy* refers to the degree to which using a technology will provide benefits to consumers in performing certain activities. In the health context, benefit examples may include managing chronic conditions or receiving health information to facilitate behavior change. *Effort expectancy* refers to the degree of ease associated with the consumers’ use of technology (ie, usability). *Social influence* refers to the extent to which consumers perceive important others, such as their family and friends, to believe they should use a particular technology. Separately, there are 2 constructs within this framework (ie, intention and facilitating conditions) that are considered direct determinants of technology usage behavior, with facilitating conditions referring to the perception of resources and support available to perform a behavior. In addition, there were 4 moderators embedded in this original framework, which contributed to understanding the acceptance of technology by individuals (ie, age, gender, experience, and voluntariness of use).

To tailor this theory to the consumer use of technology, Venkatesh et al developed UTAUT2 [44]. Given the focus of this study, it seems especially apt to use this tailored theory to understand minorities’ use of patient-centered technologies. Under this remodeled framework, 4 key constructs on the general and consumer adoption and use of technologies have been identified and incorporated into UTAUT. These additional constructs are (1) hedonic motivation (the fun or pleasure derived from using a technology), (2) price value (the monetary cost of use on the individual), (3) experience (the passage of time from initial use of the technology), and (4) habit (the extent to which an individual believes the behavior to be automatic).

Table 1. Constructs of the consumer acceptance model of the unified theory of acceptance and use of technology (UTAUT).

Constructs	Operational definitions
UTAUT constructs	
Performance expectancy	The degree to which using a technology will provide benefits to consumers in performing certain activities
Effort expectancy	The degree of ease associated with the consumers’ use of technology
Social influence	The extent to which consumers perceive the important others (family and friends) believe they should use a particular technology
UTAUT2 constructs	
Hedonic motivation	The fun or pleasure derived from using a technology
Price value	The monetary cost of use on the individual
Experience	The passage of time from initial use of the technology
Habit	The extent to which an individual believes the behavior to be automatic

Figure 1. Venkatesh et al's [44] consumer acceptance model of the unified theory of acceptance and use of technology.



Methods

Search Strategy

Recommendations of the statements on enhancing transparency in reporting the synthesis of qualitative research [45] and the preferred reporting items for systematic reviews and meta-analyses [46] were followed. Web-based search was conducted in the following academic databases: (1) PubMed (cancer subset), (2) MEDLINE, (3) PsycINFO, and (4) CINAHL. To optimize search results, we used various combinations of keywords taken from the existing literature and Medical Subject Headings terms. A complete list of search terms is provided in Table 2. Finally, we identified additional studies using a snowball searching technique whereby the reference lists of studies that met our inclusion criteria were examined.

Inclusion Criteria

We identified papers that appeared in peer-reviewed journals and were published in the English language up to October of 2016. We included both qualitative and quantitative studies and excluded nonempirical studies such as commentaries as well as international studies. Similar to other reviews [35], we limited eligible studies to those conducted in the United States because sociocultural differences in the United States may be unique from other countries. Studies were included if they assessed patient-centered technologies among underserved populations. The specific topics of interest for this review included (1) the effect of these technologies on the health or health care outcome studied; (2) the use, usability, and acceptance of these technologies and efforts to tailor their design to populations of interest; (3) facilitators and barriers to the use of patient-centered technologies among underserved populations; and (4) implementation lessons learned from studies assessing these technologies among underserved populations. Studies were included if they focused exclusively on underserved populations or underserved groups represented at least 40% of their sample size. To categorize the health information technology (HIT) apps of interest in this study, the following definitions were used.

Definitions

Electronic Health

Although no standard definition for eHealth exists [11], the term eHealth has been used broadly in the literature to refer to technologies ranging from CD-ROMs to the internet. For purposes of this study, eHealth is defined as “...the use of emerging information and communication technology, especially the internet, to improve or enable health and health care” [47].

Mobile Health

mHealth technologies are defined as “...a personalized and interactive service whose main goal is to provide ubiquitous and universal access to medical advice and information to any users at any time over a mobile platform” [48]. mHealth technologies can include the use of cell phones, smartphones, and tablets by patients or health care providers.

Telemedicine

Telemedicine has been defined as “...a branch of e-health that uses communications networks for delivery of health care services and medical education from one geographical location to another” [49]. The concept of distance is essential, for example, telemedicine can improve access to care to rural populations by eliminating distance as a barrier.

Study Selection

Each study was individually assessed for relevance. Any disagreements between reviewers were reconciled by consensus. We used a 3-step inclusion process, which is illustrated in Figure 2. In step 1, we examined paper titles and excluded papers that clearly did not have a focus on either patient-centered technologies or cancer care. In step 2, the remaining citation abstracts were retrieved. Similarly, we then excluded paper abstracts that clearly did not have a focus on either patient-centered technologies or cancer care. Finally, the full-text papers of the remaining citations were obtained for independent assessment of all inclusion criteria.

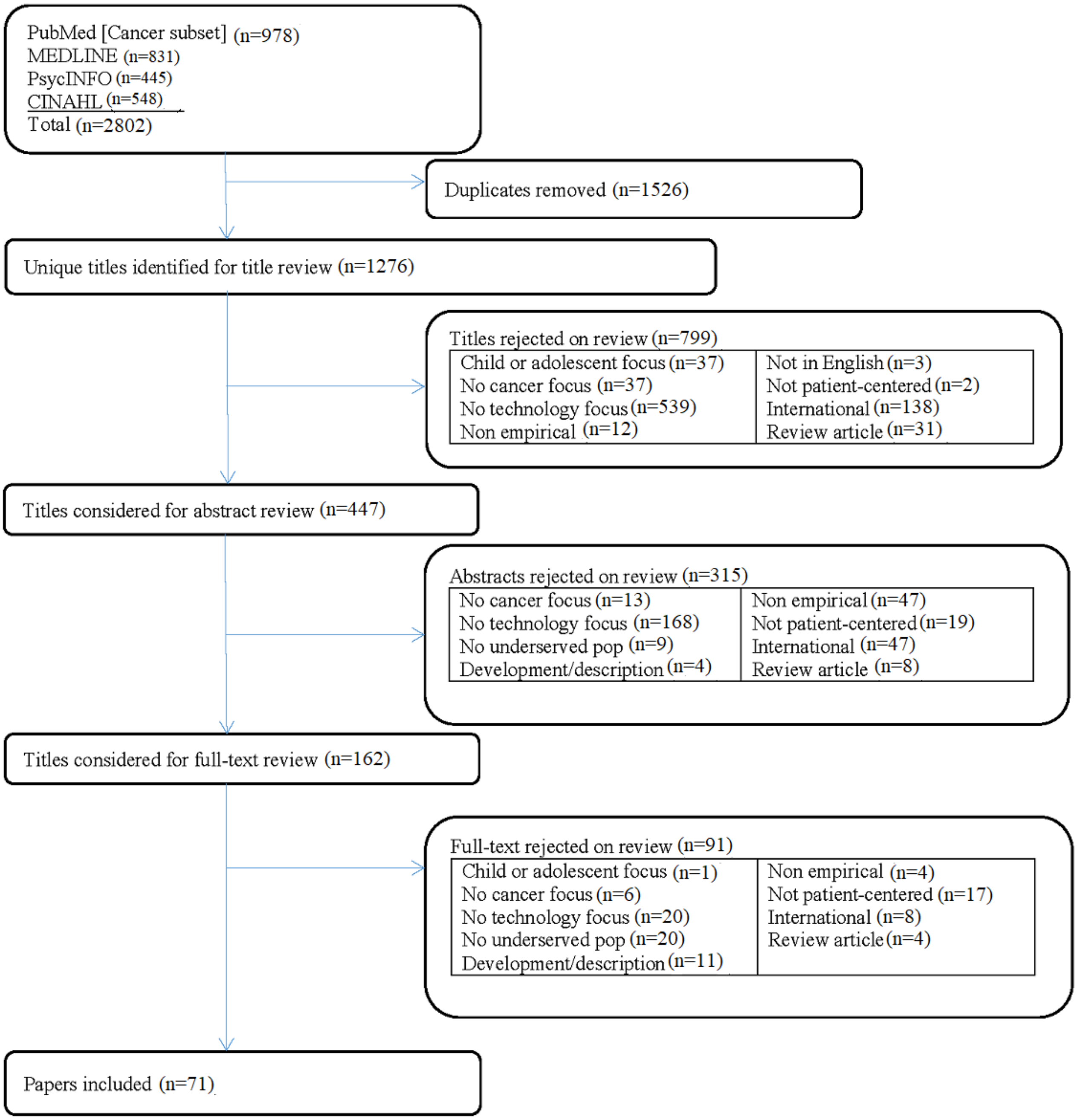
Table 2. Operationalization of the search terms.

Category	Search terms
Cancer[a]	Cancer, neoplasms
Underserved populations	Ethnic[b], race, racial, disparity[b], minority[b], underserved, rural, hispanic[b], mexican[b], latino[b], african[b], black[b], Asian, american indian[b], alaskan native[b], native american[b], inuit[b] or pacific islander[b]
Health information technology	health information technology, health it, electronic health records, electronic health record[b], electronic medical record[b], personal health record[b], personal medical record[b], patient accessible record[b], patient portal[b], patient internet portal[b], decision support[b], clinical reminder[b], electronic reminder[b], reminder system[b], m-health, mhealth, mobile technolog[b], mobile health, cell phone[b], cellular phone[b], smartphone[b], mobile phone[b], mobile device[b], text message[b], cd-rom, dvd, computer based, computer-based, internet-based, web-based, web based, e-health, ehealth, tablet, tailored, telemedicine, telehealth, teleoncology

[a] Search terms within each category are combined with OR. Search terms between categories are combined with AND. Some terms were truncated.

[b] Truncation of search term to capture keywords with the same stem.

Figure 2. Systematic review flowchart. pop: population.



Data Extraction

Information systematically extracted from the papers included the following: study design, including the targeted cancer and/or stage of the cancer care continuum [50]. The continuum of cancer care refers to prevention, detection, diagnosis, treatment, survivorship, and end-of-life care [51]. In addition, we extracted information on the underserved population of interest; whether the patient-centered technology focused on healthy individuals, cancer patients or survivors, caregivers, or health care providers; sample size; the type of patient-centered technology used; the study outcome of interest; and whether there was any evidence of tailoring when it came to the technology intervention. Tailoring is defined as “Any combination of information or change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and have been derived from an individual assessment” [52]. Moreover, when available we described the technology’s use (whether patients adopt the tool), usability (the patient’s experience using the tool), and usefulness (the extent to which it meets the patient’s needs). Finally, we described barriers and facilitators to the use of the technologies reported in the paper.

Results

Studies Included

Our keyword search identified an initial yield of 1276 nonduplicative studies (Figure 2). The primary reasons for exclusion are also identified in Figure 2. After applying the exclusion criteria in review of the titles and abstracts, 71 studies were included in the systematic review (marked with an asterisk in the reference list), published between 1995 and 2016.

Study Characteristics

The characteristics of the studies are summarized in Table 3. Studies varied with respect to the underserved population targeted, the technology used, and the cancer type of interest. A large proportion of studies included in our review target blacks or African Americans (31/71, 44%) followed by rural populations (14/71, 20%). More than half of the included studies assessed eHealth technologies (41/71, 58%). Moreover, the largest proportion of studies focused on breast cancer (26/71, 37%). In addition, the largest proportion of technological outcomes assessed was use of technology (20/71, 2%), whereas knowledge (15/71, 21%) was the largest proportion of health outcomes assessed. Observational studies represented the largest proportion of studies included in our review (32/71, 45%). Overall, 15 studies followed an experimental design (15/71, 21%), whereas the remaining 24 studies were either qualitative (13/71, 18%) or mixed-methods (11/71, 16%).

To provide a consistent structure, the remainder of the Results section is organized as follows. Study summaries are stratified by the type of patient-centered technology. Within each patient-centered technology section, we then further stratify by study design: experimental (Multimedia Appendix 1), observational (Multimedia Appendix 2), and qualitative studies (Multimedia Appendix 3).

Theme 1. The Effect of Use on Clinical and Other Outcomes

Electronic Health

In total, 10 eHealth studies were identified using an experimental design, with 9 of these using a randomized controlled trial (RCT) [41,53-61] (Multimedia Appendix 1). These studies primarily targeted African Americans (n=7) as well as colorectal (n=6) and breast (n=4) cancer. Findings showed positive impacts of eHealth interventions, ranging from computer-assisted programs and Web-based decision aids to tailored, interactive soap operas. The most common health outcomes assessed included knowledge (n=5) and screening uptake (n=3), with each study showing statistically significant results in the outcomes.

For example, Champion et al [41] used an RCT to compare the efficacy of 3 interventions in promoting routine mammography screening among low-income African American women. This study found that an interactive computer-assisted instruction program produced the greatest adherence to mammography (40.0%) compared with participants receiving an educational pamphlet (32.1%) or a culturally appropriate video (24.6%). Jibaja et al [57] used an interactive soap-opera format to promote the early detection of breast cancer among high-risk Hispanic women. The use of this culturally tailored, computer-based educational program was found to significantly increase breast cancer screening knowledge and beliefs relative to a comparison group.

Mobile Health

We identified 3 mHealth studies using an experimental design [62-64], with 2 studies (66.7%) showing statistically significant results in the health outcome. Using a quasi-experimental design, Lee et al [64] tested a tailored interactive 7-day short message service (SMS) text message intervention designed to increase knowledge and vaccination of human papillomavirus (HPV). This study found a significant increase in knowledge and intent to get vaccinated. In addition, HPV vaccination uptake increased by 30% among participants in the intervention. Targeting a Hispanic population for CRC screening, Fernandez et al [63] used an RCT to compare (1) a tailored interactive multimedia intervention, (2) a lay health worker delivered media print intervention, and (3) a no intervention control group. No statistically significant differences were found among the study arms. Among a population with advanced prostate cancer that included a significant proportion of African Americans (40.5%), Yanez et al [62] found that a Web-based psychosocial intervention delivered via a tablet achieved good retention (>85%) and attendance rates (>70%) and received favorable evaluations (mean score: 4/5) and exit surveys (mean score: 3.6/4). The intervention also reduced depressive symptoms (43.37 vs 47.29, P=.03) and improved relaxation self-efficacy (2.43 vs 1.11, P<.01) for men who completed the study.

We also identified 2 additional studies using a mixed-methods approach with an experimental component. Among a sample of Korean American women, a 7-day mobile phone text message–based cervical cancer screening intervention significantly increased participants' knowledge of cervical cancer and screening recommendations as well as the uptake of cervical cancer screening [42]. In a sample of Spanish-speaking Latina women seeking care at a federally qualified health center, text messaging reduced the number of days between an abnormal mammogram and participants' return for follow-up compared with women who did not receive text message notifications [65].

Telemedicine

We identified 2 telemedicine studies using an experimental design, with each study showing statistically significant results in the health outcomes [66,67]. Kroenke et al [67], using an RCT design, found that telecare management improved depression and pain outcomes in both urban and rural cancer patients. In another RCT, telegenetics was compared with in-person cancer genetic counseling in terms of its impact on attendance, patient satisfaction, and cost [66]. This study found that although costs were significantly less, telegenetics did not differ in patient satisfaction from in-person genetic counseling; however, patients seeking in-person genetic counseling were more likely to attend counseling sessions.

Table 3. Characteristics of studies included in this review (N=71).

Characteristics	Total, n (%)
Underserved population	
American Indian or Alaskan native	3 (4)
Asian	6 (8)
Black or African American	31 (44)
Hispanic	12 (17)
Diverse pop	4 (6)
Low income	6 (9)
Rural	14 (20)
Patient-centered technology	
Computer- or internet-based technology (eHealth)	41 (58)
Mobile app (mHealth)	15 (21)
eHealth and mHealth	5 (7)
Telemedicine	10 (14)
Cancer type	
Breast	26 (37)
Cervical	4 (6)
Colorectal	12 (17)
Lung	1 (1)
Ovarian	1 (1)
Prostate	9 (13)
Cancer (not specific)	18 (25)
Technology outcomes	
Use	20 (28)
Usefulness	5 (7)
Usability or acceptability	18 (25)
Design or implementation	6 (9)
Satisfaction	5 (7)
Health outcomes	
Communication	1 (1)
Decision making	4 (6)
Health beliefs	2 (3)
Intention or readiness	2 (3)
Knowledge	15 (21)
Participation in health care	2 (3)
Pain	1 (1)
Psychological	5 (7)
Quality of life	2 (3)
Satisfaction	1 (1)
Vaccination	1 (1)
Screening	10 (14)
Study design	
Experimental	15 (21)

Characteristics	Total, n (%)
Observational	32 (45)
Qualitative	13 (18)
Mixed methods	11 (16)

Theme 2. Behavioral Intention to Use, Use, and Usefulness of Patient-Centered Technology

Electronic Health

Overall, 17 observational studies were identified for eHealth, with the majority targeting African Americans (n=9) (Multimedia Appendix 2) [68-84]. Most studies assessed cancer in general (n=8), with breast cancer being the most frequent single site focus (n=6). Studies found that the internet was the first source of cancer information, followed by health care providers, for Hispanics [68] and blacks [75]. In addition, email and Web-based information was preferred over mail [78]. Among a low-income population, Song et al [82] found that the internet was identified as the least relied upon source of general health information and cancer health information compared with *health professionals, family, and friends*. A majority of studies assessed technology use as an outcome (n=15).

In an early study, Gustafson et al [73] examined the feasibility of an interactive, computer-based system in reaching low-income, underserved women with breast cancer (N=229; n=85 African Americans) [85]. Low-income women were more likely to use and spend more time on the computer-based system compared with another population of more affluent women. In addition, low-income urban African Americans were more likely to use the system to access information and for health management services, whereas low-income whites were more likely to use communication services. The *Young Sisters Initiative: A Guide to a Better You!* program is a website designed for young breast cancer survivors. Using a mixed-methods approach, which included a postuse survey of 1442-site visitors (93% African American women), participants reportedly found value in using the website for reproductive and psychosocial information and support [86]. Chee et al [87] conducted a usability test and RCT pilot intervention to determine the efficacy of a culturally tailored registered nurse-moderated internet cancer support group. This study found positive effects on supportive care needs, psychological and physical symptoms, and quality of life.

In addition to personal computer use, the use of computer kiosks was explored among underserved populations. Kreuter et al [69] sought to understand the ideal placement (eg, beauty salons, churches, neighborhood health centers, laundromats, social service agencies, health fairs, and public libraries) to reach African American women for the purpose of providing tailored breast cancer information. This study found that only laundromats resulted in both frequent kiosk use and reaching high need populations (ie, a large proportion of users with no health insurance, unaware of where to get a mammogram, reporting no recent mammogram and barriers to getting one, and having little knowledge about breast cancer and mammography).

Mobile Health

We found 7 observational studies assessing mHealth (n=2) [88,89] or both mHealth and eHealth studies (n=5) [36-40]. Surveying 156 Hispanic and non-Hispanic rural women, Kratzke and Wilson [37,88] found that nearly 87% of study participants used cell phones, whereas 47% used text messaging as a means to communicate. Compared with non-Hispanic women, Hispanic women (n=36) were more receptive to breast cancer prevention voice messages and text messages. In another survey of Hispanic women (n=905), Dang et al [36] found that more than half of participants did not use the internet (58%) or email (64%), but a large proportion of participants used mobile phones (70%). In addition, 65% of all participants used text messages, with 45% wishing to receive mammogram reminders via SMS text message. Schoenberger et al used focus groups to assess the usage and acceptance of mobile communication technologies to provide cancer information among community health advisors (n=37) [90] and health ministry leaders (n=37) [91]. Among community health advisors, a majority of participants reported owning a mobile phone (89%) or a smartphone (67%) and 33% use text messaging as a means to communicate. All health ministry leaders reported owning a phone, whereas 85% reported using text messaging as a means to communicate.

Telemedicine

We identified 8 telemedicine studies using an observational design [92-99]. Of these, 3 studies focused on genetic counseling. Using surveys, McDonald et al [97] sought to understand the acceptability of telegenetics among Maine residents living in rurally remote areas. The most important characteristics of telegenetics models of care were perceived to be professional qualifications (92.2%) and one-on-one counseling (65.1%), whereas in-person and local counseling was ranked lower (51.8% and 52.1%, respectively).

Telemedicine was commonly used to provide psychosocial support. Rural lung, breast, and colorectal cancer patients reported a high level of satisfaction with a videophone-based intervention providing dignity psychotherapy [98]. Among a sample of Alaskan native breast cancer patients, an interactive audio and video telemedicine program providing medical consultation received overall high patient satisfaction [99]. In a sample of rural American Indian and Alaskan natives in Washington, cancer survivors were surveyed about their experiences with a telehealth cancer support group [93]. Members reported value in interacting with other cancer survivors and usefulness of the information presented. Specific topics of interest included nutrition during treatment as well as side effects of treatment.

Theme 3. Perceptions and Satisfaction of Use of Patient-Centered Technologies

Mobile Health

We identified 5 qualitative studies related to mHealth (Multimedia Appendix 3) [91,100-103]. Qualitative studies primarily targeted black or African American (n=2) or Hispanic (n=2) populations and breast cancer (n=3) patients. These studies assessed outcomes related to content design and implementation (n=3) and usability or acceptability (n=2). Weaver et al [104] used focus groups to assess the perceptions of colorectal cancer screening text messages among a majority African American population (n=16; 62%). Although initially expressing reluctance to use personal technologies as a means to receive CRC information, participants responded favorably when shown sample text messages. Features that participants were interested in seeing with respect to text messages were personalized messages, content that was relevant to them, and messages that were positive and reassuring. Conversely, participants did not want to receive test results or bad news via text messages or content that included shorthand phrases or required complex replies. In a group of healthy African American men who received a prostate cancer screening educational intervention consisting of short text messages related to prostate cancer awareness, Le et al [105] found that 65% of the participants wished to continue receiving text messages pertaining to workshop reminders, postworkshop reinforcement, spiritual or motivational messages, and retention after completing the study.

Bravo et al [100] used semistructured interviews to assess the attitudes, acceptance, and usability of a breast cancer risk assessment tool accessed via tablet among underserved women seeking care at a safety net institution. A majority of women preferred the mobile app over a paper version of the assessment tool. All participants found the app easy to use.

Electronic Health

We identified 7 qualitative studies [106-112] and 7 mixed methods [86,87,113-117] related to eHealth. Qualitative studies primarily targeted black or African Americans (n=3) or diverse populations (n=2) and targeted breast (n=3) and prostate cancer (n=3) patients. These studies largely assessed outcomes related to usability or acceptability (n=3) and content design (n=2). For example, Berry et al [106] evaluated the usability of a Web-based decision aid designed to improve decision making among English-speaking Latino men (n=7) with localized prostate cancer. This eHealth intervention was tailored to participants' personal factors (eg, personal characteristics, confidence in doctor, and influential people) and used expert recommendations to communicate health benefits and risks. Overall, participants rated the intervention with high acceptability. However, Berry et al found several usability issues related to *content comprehension* where Hispanic participants did not initially understand concepts until provided a short definition; *navigation issues* when answering multiple choice questions, using check boxes, typing responses with a keyboard, or clicking links to access external pages; and *sociocultural appropriateness* where some subgroups of the population (eg, Latino men in poverty) did not have computers at home and would therefore not use the app.

Theme 4. Barriers and Facilitators to Use of Patient-Centered Technologies

Barriers

Text messaging was found to be beneficial as it provided a form of communication allowing for the quick dissemination of health information. The main barrier to text messaging was the lack of knowledge of how to text among some participants. [90,91] Increasing knowledge through education was found to be the most feasible solution. Another barrier to successful outcomes from text messaging [105] was read or receipt, wherein some participants did not remember receiving text messages. This problem could have arisen from technical issues and suggested the importance of incorporating a component into the intervention that verified that messages had been received.

Facilitators

Participants considered a complex mobile app to be easy and fun to use with easy to read text. In addition, participants were motivated to use the tool as it was new and innovative [100]. Although participants reported being unfamiliar with the technology (iPad) and experienced challenges with health literacy and security concerns, they had an interest in keeping up with technology [100]. Some suggestions identified when designing mHealth apps for this population include writing the content at the appropriate literacy levels, making instruction and assistance in using the mHealth app available, and minimizing the amount of new skillsets that participants will need to learn to use the mHealth app.

Discussion

Principal Findings

The purpose of this study was to review the current evidence on the use of cancer-specific patient-centered technologies among the underserved. Although the reviewed studies targeted various underserved populations including racial and ethnic minorities (eg, blacks or African Americans and Hispanics), low-income, and rural populations, we identified 2 cross-cutting issues that the literature suggests should be taken into account when implementing patient-centered technology interventions: (1) training in the use of patient-centered technologies and (2) tailoring patient-centered technologies to target populations.

The landscape of technology in our digital age is rapidly changing. This growth has led to several advances in health promotion from accessing health information digitally to using technology to track health and fitness [118]. In addition, the internet and mobile devices have become a prominent vehicle to reach diverse minority populations and deliver health information [23-28]. Use of the internet within the home is lower in individuals who are older, belong to a racial or ethnic minority group, are less educated, and have lower incomes [119,120]; however, the internet has become more accessible in many ways because of the proliferation of mobile devices. For example, blacks are more likely to access the internet with their mobile phone than their non-Hispanic white counterparts [119].

Much of the evidence related to mobile devices was devoted to the use of text messaging as a means to provide health information and facilitating behavior change. These results are promising given the consistent findings that underserved populations are receptive to the use of these technologies for cancer prevention and care purposes. Contributing factors to this growth, and the intention to use these technologies, may be traced back to the constructs of *social influence*, *price value*, and *habit*. According to the Pew Research Center, text messaging is being used by more than 90% of the population within each age group (100% of 18-29 years, 98% of 30-49 years, and 92% of >50 years) [121]. Due to this widespread use and already established habit of communicating with others, individuals are likely to adopt this form of technology to communicate with their social group. Due to the way in which this method of communication is ingrained into the day-to-day lives of individuals, the time cost of adopting these technologies is minimal because of their pervasiveness. In addition, using smartphones may be considered a low-cost alternative to accessing the internet compared with home internet.

Mobile devices provide a means to reach minority populations and offer the potential to reduce access issues with respect to health care and health information. However, barriers still exist that prevent the effective use of these technologies. In addition to creating opportunities to advance health promotion, the rapid growth in technology also presents several challenges. Of the most prominent challenges facing users, 1 is the pressure to remain updated with new technologies, their increasing *effort expectancy*, or the degree of ease associated with technology use. Clearly, realizing the full potential benefit of these technologies is dependent on their effective use. Although studies found that underserved populations are receptive to the use of patient-centered technologies, a recurrent challenge found in the literature was a lack of knowledge as to how to use new apps of the technology as well as the technology itself. These challenges should not be overlooked and range from receiving health information via text messages to using interactive iPads.

Education and Training to Facilitate the Use of Patient-Centered Technologies

Some of these difficulties with use could be remedied by *facilitating conditions*, for example, providing a short training session at the same time the technology is introduced. Public libraries have been successful in improving decision making in accessing high-quality health information, reducing computer anxiety, and increasing computer interest and self-efficacy among older adults [122,123], and health care providers could leverage or learn from these community institutions. However, other technologies may require substantial modifications to the intervention to remove obstacles and barriers individuals may experience to facilitate their use. Interventions should incorporate usability and feasibility testing with target populations into their development process to identify unanticipated issues as well as appropriate training of target populations in the use of the technologies. Although such methods need to be applied efficiently to minimize their time and resource burden, up-front investment in such approaches can be the difference between a successful or failed implementation. In some cases, new technologies may not be a good fit with underserved populations, for example, low-income individuals may have insurmountable barriers to obtaining expensive new devices. Patients with disabilities, whether mental or physical, may not have the capacity to adapt to new technologies that require significant cognitive load or fine motor skills. In these cases, alternative communication channels may be necessary to deliver a desired behavioral or clinical intervention; we want to be careful to construct patient-centered technologies versus technology-centered patients.

Tailoring to Facilitate the Use of Patient-Centered Technologies

When using interactive technologies (ie, computer-based media that enable users to access information and services of interest, control how the information is presented, and respond to information and messages in the mediated environment [124]), an important feature is the ability to tailor information to the recipients' needs and interests [124,125]. It is necessary to consider the unique cultural norms and/or challenges of underserved populations when tailoring communication strategies. Robust methods to account for these differences in the design and implementation of technology interventions targeting specific groups is a key area in need of development. Hispanics may be better reached with technologies framed with health education content tailored to this population to improve both content comprehension and acceptance. For example, the use of telenovelas and soap operas is a novel approach that appeals to underserved Spanish-speaking women's cultural norms and has been found to increase breast cancer screening knowledge and beliefs [57]. Similarly, a culturally tailored educational video, including a soap opera and physician recommendation segment made in Chinese was found to increase Chinese women's intention to get screened for breast cancer, in addition to increasing their knowledge, perceived risk, and perceived benefits of screening [116].

Cultural competence is another strategy to reduce health and health care disparities that may be applied to the tailoring of patient-centered technologies. Cultural competence is defined as "...understanding the importance of social and cultural influences on patients' health beliefs and behaviors; considering how these factors interact at multiple levels of the health care delivery system (eg, at the level of structural processes of care or clinical decision making); and, finally, devising interventions that take these issues into account to assure quality health care delivery to diverse patient populations" [126]. Tailoring patient-centered technologies to patients may help overcome sociocultural barriers to providing health care, one being a lack of culturally or linguistically appropriate health education materials [126]. By understanding unique differences among underserved groups, we can better understand how to reach each population, how they spend their time and use technology, and how different forms of technology may be used in different home and community settings. In addition, this approach allows researchers to tailor the technology based on who an individual is and how their identity is constructed before the technology is implemented.

Directions for Future Research

Digital Divide

The use of patient-centered technologies may be seen as a means to reach underserved populations; however, there are several concerns within the health care research community related to their use. Of particular interest is the decreased access of technologies among racial and ethnic minorities, persons with disabilities, rural populations, older populations (including veterans), and individuals with lower socioeconomic status; a phenomenon commonly referred to as the digital divide [127,128]. More research needs to assess the health information needs of these underserved populations and how they prefer to receive health information. When the use of technology may not be appropriate for providing patient-centered care, the use of other tailored interventions may be more successful.

Underrepresented Cancers and Underserved Populations

According to the American Cancer Society, the most prevalent cancer among men is prostate cancer, whereas breast cancer is the most common cancer among women [129]. In addition, lung and colorectal cancer comprise the second and third most common cancer in both men and women [129]. Although several studies have targeted breast cancer, colorectal, and prostate cancer, only a few studies have targeted other cancers. Future research on the use of patient-centered technologies among underserved populations should focus on prevalent cancers, which are underrepresented in the HIT literature such as lung cancer. Importantly, lung cancer is the leading cause of cancer deaths in black men and women, and black men have higher rates of lung cancer than their non-Hispanic white counterparts [130]. Furthermore, less prevalent cancers have not seen the same level of technology development targeting their unique clinical needs. Although some cancers are less prevalent in the general population, they may disproportionately impact underserved populations. For example, Hispanic men experience liver cancer incidence rates twice that of non-Hispanic white men [131]. Liver cancer also serves as the second leading cause of cancer deaths. Cancers of the female reproductive system are also underrepresented, including ovarian and cervical cancers. African American women are diagnosed with more advanced stages of ovarian cancer and have lower survival rates compared with their white counterparts. In addition, there are underserved populations that have received little attention in the current literature. Studies predominantly targeted black or African Americans and Hispanic populations, and some underrepresented populations include American Indians or Alaskan Natives and Asian populations.

Underused Technologies

The current evidence with respect to barriers and facilitators to the use of patient-centered technologies may be used to guide the development of other technologies, such as PHRs, which did not appear in our review. PHRs have been defined as “an Internet-based set of tools that allows people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it.” PHRs can be tethered (connected) to eHealth records and provide patients with an asynchronous platform to access and update their medical health record data and engage with their health care team [9,132,133]. As an example, the unique challenges we identified among Hispanics, including issues related to tailored educational content and comprehension, may be used to modify patient portals within practices serving a Hispanic community.

Patient-Provider Communication and Shared Decision Making

Our review also highlights the paucity of research regarding how health information technology can improve communication and shared decision making (SDM) between individuals from vulnerable populations and their health care providers. Although increasing provider communication is important in building trust and improving chronic disease management [4], SDM bridges gaps in knowledge, tailors medical and health decisions to patient preferences, as well as increases patient adherence to treatment and improves health outcomes [134]. More research should focus on the use of these technologies to support providers in delivering information to patients on cancer treatment options as well as describe the advantages and disadvantages of different approaches to technology design and implementation.

Precision Medicine

By tailoring content to targeted populations, patient-centered technologies have the potential to facilitate the provision of precision medicine among underserved populations. Precision medicine focuses on the “...prevention and treatment strategies that take individual variability into account” [135]. Interactive technologies such as social media may take advantage of predictive algorithms to tailor the care of individuals to patients based not only on their genetic but also on their social identities.

Limitations

Our study has several limitations. Of the potential limitations, 1 is that our search strategy may not have captured all potential papers meeting our inclusion criteria. To minimize this limitation, we implemented a snowball search method in which we reviewed the references of all included studies for additional citations. Another limitation of our study is that because of the heterogeneity in study design and types of outcomes evaluated, we were unable to aggregate findings in the manner of a meta-analysis. Finally, the included papers may be subject to publication bias as studies that report negative findings are less likely to be published.

Conclusions

There is a growing body of literature exploring patient-centered technology and its influence on the care of underserved populations. Despite the potential of patient-centered technologies and their acceptance among underserved populations, challenges still exist with respect to their effective use and usability. With technology changing at an exceedingly rapid pace, more training needs to be provided to ensure these underserved groups are able to effectively use new and emerging technologies. In addition, tailoring these technologies to unique cultural norms will be critical to facilitating their effective use.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description and findings of experimental studies included in this review.

[PDF File (Adobe PDF File), 93KB - jmir_v21i4e10256_app1.pdf]

Multimedia Appendix 2

Description and findings of observational studies included in this review.

[PDF File (Adobe PDF File), 114KB - jmir_v21i4e10256_app2.pdf]

Multimedia Appendix 3

Description and findings of qualitative and mixed-methods studies included in this review.

[PDF File (Adobe PDF File), 93KB - jmir_v21i4e10256_app3.pdf]

References

1. 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]
2. Health Resources & Services Administration. 2016. Medically underserved areas and populations (MUA/Ps) URL: https://bhw.hrsa.gov/shortage-designation/muap [accessed 2019-02-26] [WebCite Cache ID 76UR0VeYy]
3. American Cancer Society. 2011. Special section: Cancer disparities and premature deaths. Cancer facts & figures 2011 URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2011.html [accessed 2019-02-27] [WebCite Cache ID 76VupgcIu]
4. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academy Press; 2001.
5. Atrash HK, Hunter MD. Health disparities in the United States: a continuing challenge. In: Satcher D, Pamies RJ, Woelfl NN, editors. Multicultural Medicine and Health Disparities. New York: McGraw-Hill; 2006.
6. Smedley B, Stith A, Nelson A. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academy Press; 2003.
7. Gibbons MC, Fleisher L, Slamon RE, Bass S, Kandadai V, Beck JR. Exploring the potential of Web 2.0 to address health disparities. J Health Commun 2011;16(Suppl 1):77-89. [doi: 10.1080/10810730.2011.596916] [Medline: 21843097]
8. López L, Green AR, Tan-McGrory A, King R, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011 Oct;37(10):437-445. [doi: 10.1016/S1553-7250(11)37055-9] [Medline: 22013816]
9. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
10. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
11. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
12. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
13. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
14. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
15. Ueckert F, Goerz M, Ataian M, Tessmann S, Prokosch H. Empowerment of patients and communication with health care professionals through an electronic health record. Int J Med Inform 2003 Jul;70(2-3):99-108. [doi: 10.1016/S1386-5056(03)00052-2] [Medline: 12909161]

16. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
17. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [doi: 10.1016/S1386-5056(00)00130-1] [Medline: 11248599]
18. Kupchunas WR. Personal health record: new opportunity for patient education. Orthop Nurs 2007 Mar;26(3):185-91; quiz 192. [doi: 10.1097/01.NOR.0000276971.86937.c4] [Medline: 17538475]
19. 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]
20. Custodio R, Gard AM, Graham G. Health information technology: addressing health disparity by improving quality, increasing access, and developing workforce. J Health Care Poor Underserved 2009 May;20(2):301-307. [doi: 10.1353/hpu.0.0142] [Medline: 19395831]
21. Christopher GM. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011;8:1f [FREE Full text] [Medline: 21307989]
22. US Department of Health & Human Services: Office of Minority Health. 2010. HHS action plan to reduce racial and ethnic health disparities: A nation free of disparities in health and health care URL: https://minorityhealth.hhs.gov/npa/files/plans/hhs/hhs_plan_complete.pdf [accessed 2019-02-26] [WebCite Cache ID 6xWiiWjru]
23. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]
24. Institute of Medicine. Speaking of Health: Assessing Health Communication Strategies for Diverse Populations. Washington, DC: National Academies Press; 2002.
25. Birru M, Steinman RA. Online health information and low-literacy African Americans. J Med Internet Res 2004 Sep 3;6(3):e26 [FREE Full text] [doi: 10.2196/jmir.6.3.e26] [Medline: 15471752]
26. Jackson CL, Batts-Turner ML, Falb MD, Yeh H, Brancati FL, Gary TL. Computer and internet use among urban African Americans with type 2 diabetes. J Urban Health 2005 Dec;82(4):575-583 [FREE Full text] [doi: 10.1093/jurban/jti126] [Medline: 16221917]
27. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
28. Viswanath K, McCloud R, Minsky S, Puleo E, Kontos E, Bigman-Galimore C, et al. Internet use, browsing, and the urban poor: implications for cancer control. J Natl Cancer Inst Monogr 2013 Dec;2013(47):199-205 [FREE Full text] [doi: 10.1093/jncimonographs/lgt029] [Medline: 24395992]
29. Kim K, Nahm E. Benefits of and barriers to the use of personal health records (PHR) for health management among adults. Online J Nurse Inform 2012;16(3) [FREE Full text]
30. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
31. 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]
32. Apter AJ. Can patient portals reduce health disparities? A perspective from asthma. Ann Am Thorac Soc 2014 May;11(4):608-612. [doi: 10.1513/AnnalsATS.201401-032PS] [Medline: 24640983]
33. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
34. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014 Oct;15(Suppl 4):146-158. [doi: 10.1111/obr.12218] [Medline: 25196411]
35. Montague E, Perchonok J. Health and wellness technology use by historically underserved health consumers: systematic review. J Med Internet Res 2012 May 31;14(3):e78. [doi: 10.2196/jmir.2095] [Medline: 22652979]
36. Dang CM, Estrada S, Bresee C, Phillips EH. Exploring potential use of internet, e-mail, and instant text messaging to promote breast health and mammogram use among immigrant Hispanic women in Los Angeles County. Am Surg 2013 Oct;79(10):997-1000. [Medline: 24160786]
37. Kratzke C, Wilson S, Vilchis H. Reaching rural women: breast cancer prevention information seeking behaviors and interest in internet, cell phone, and text use. J Community Health 2013 Feb;38(1):54-61. [doi: 10.1007/s10900-012-9579-3] [Medline: 22706805]
38. Haughton LT, Kreuter M, Hall J, Holt CL, Wheetley E. Digital divide and stability of access in African American women visiting urban public health centers. J Health Care Poor Underserved 2005 May;16(2):362-374. [doi: 10.1353/hpu.2005.0034] [Medline: 15937398]

39. Kratzke C, Amatya A, Vilchis H. Differences among college women for breast cancer prevention acquired information-seeking, desired apps and texts, and daughter-initiated information to mothers. J Community Health 2014 Apr;39(2):291-300. [doi: 10.1007/s10900-013-9759-9] [Medline: 23979671]
40. Robinson RF, Dillard DA, Hiratsuka VY, Smith JJ, Tierney S, Avey JP, et al. Formative evaluation to assess communication technology access and health communication preferences of Alaska Native people. Int J Indig Health 2015;10(2):88-101 [FREE Full text] [doi: 10.18357/ijih.102201515042] [Medline: 27169131]
41. Champion VL, Springston JK, Zollinger TW, Saywell RM, Monahan PO, Zhao Q, et al. Comparison of three interventions to increase mammography screening in low income African American women. Cancer Detect Prev 2006;30(6):535-544. [doi: 10.1016/j.cdp.2006.10.003] [Medline: 17110056]
42. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
43. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
44. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
45. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
46. 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]
47. Eng TR. The Ehealth Landscape: A Terrain Map of Emerging Information and Communication Technologies in Health and Health Care. Princeton, NJ: Robert Wood Johnson Foundation; 2001.
48. Akter S, Ray P. mHealth - an ultimate platform to serve the unserved. Yearb Med Inform 2010:94-100. [doi: 10.1055/s-0038-1638697] [Medline: 20938579]
49. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
50. National Cancer Institute. 2017. Cancer control continuum URL: https://cancercontrol.cancer.gov/od/continuum.html [accessed 2019-02-26] [WebCite Cache ID 6xWjMQFmc]
51. National Cancer Institute. 2017. NCI Dictionary of Cancer Terms: Continuum of care URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/continuum-of-care [accessed 2019-02-26] [WebCite Cache ID 6xWjPsQFD]
52. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [doi: 10.1093/her/15.1.1] [Medline: 10788196]
53. Christy SM, Perkins SM, Tong Y, Krier C, Champion VL, Skinner CS, et al. Promoting colorectal cancer screening discussion: a randomized controlled trial. Am J Prev Med 2013 Apr;44(4):325-329 [FREE Full text] [doi: 10.1016/j.amepre.2012.11.032] [Medline: 23498096]
54. Ellison GL, Weinrich SP, Lou M, Xu H, Powell IJ, Baquet CR. A randomized trial comparing web-based decision aids on prostate cancer knowledge for African-American men. J Natl Med Assoc 2008 Oct;100(10):1139-1145 [FREE Full text] [doi: 10.1016/S0027-9684(15)31481-4] [Medline: 18942274]
55. Greiner KA, Daley CM, Epp A, James A, Yeh HW, Geana M, et al. Implementation intentions and colorectal screening: a randomized trial in safety-net clinics. Am J Prev Med 2014 Dec;47(6):703-714 [FREE Full text] [doi: 10.1016/j.amepre.2014.08.005] [Medline: 25455115]
56. Gustafson DH, Hawkins R, Pingree S, McTavish F, Arora NK, Mendenhall J, et al. Effect of computer support on younger women with breast cancer. J Gen Intern Med 2001 Jul;16(7):435-445 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016007435.x] [Medline: 11520380]
57. Jibaja ML, Kingery P, Neff NE, Smith Q, Bowman J, Holcomb JD. Tailored, interactive soap operas for breast cancer education of high-risk Hispanic women. J Cancer Educ 2000;15(4):237-242. [doi: 10.1080/08858190009528705] [Medline: 11199243]
58. Miller DP, Spangler JG, Case LD, Goff DC, Singh S, Pignone MP. Effectiveness of a web-based colorectal cancer screening patient decision aid: a randomized controlled trial in a mixed-literacy population. Am J Prev Med 2011 Jun;40(6):608-615 [FREE Full text] [doi: 10.1016/j.amepre.2011.02.019] [Medline: 21565651]
59. Rawl SM, Skinner CS, Perkins SM, Springston J, Wang HL, Russell KM, et al. Computer-delivered tailored intervention improves colon cancer screening knowledge and health beliefs of African-Americans. Health Educ Res 2012 Oct 1;27(5):868-885 [FREE Full text] [doi: 10.1093/her/cys094] [Medline: 22926008]
60. Russell KM, Champion VL, Monahan PO, Millon-Underwood S, Zhao Q, Spacey N, et al. Randomized trial of a lay health advisor and computer intervention to increase mammography screening in African American women. Cancer Epidemiol Biomarkers Prev 2010 Jan;19(1):201-210 [FREE Full text] [doi: 10.1158/1055-9965.EPI-09-0569] [Medline: 20056639]

61. Schroy PC, Emmons K, Peters E, Glick JT, Robinson PA, Lydotes MA, et al. The impact of a novel computer-based decision aid on shared decision making for colorectal cancer screening: a randomized trial. Med Decis Making 2011;31(1):93-107 [FREE Full text] [doi: 10.1177/0272989X10369007] [Medline: 20484090]
62. Yanez B, McGinty HL, Mohr DC, Begale MJ, Dahn JR, Flury SC, et al. Feasibility, acceptability, and preliminary efficacy of a technology-assisted psychosocial intervention for racially diverse men with advanced prostate cancer. Cancer 2015 Dec 15;121(24):4407-4415 [FREE Full text] [doi: 10.1002/cncr.29658] [Medline: 26348661]
63. Fernández ME, Savas LS, Carmack CC, Chan W, Lairson DR, Byrd TL, et al. A randomized controlled trial of two interventions to increase colorectal cancer screening among Hispanics on the Texas-Mexico border. Cancer Causes Control 2015 Jan;26(1):1-10. [doi: 10.1007/s10552-014-0472-5] [Medline: 25466604]
64. Lee HY, Koopmeiners JS, McHugh J, Raveis VH, Ahluwalia JS. mHealth pilot study: text messaging intervention to promote HPV vaccination. Am J Health Behav 2016 Jan;40(1):67-76 [FREE Full text] [doi: 10.5993/AJHB.40.1.8] [Medline: 26685815]
65. Oakley-Girvan I, Londono C, Canchola A, Watkins DS. Text messaging may improve abnormal mammogram follow-up in Latinas. Oncol Nurs Forum 2016 Jan;43(1):36-43. [doi: 10.1188/16.ONF.36-43] [Medline: 26679443]
66. Buchanan AH, Datta SK, Skinner CS, Hollowell GP, Beresford HF, Freeland T, et al. Randomized trial of telegenetics vs in-person cancer genetic counseling: cost, patient satisfaction and attendance. J Genet Couns 2015 Dec;24(6):961-970 [FREE Full text] [doi: 10.1007/s10897-015-9836-6] [Medline: 25833335]
67. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. J Am Med Assoc 2010 Jul 14;304(2):163-171 [FREE Full text] [doi: 10.1001/jama.2010.944] [Medline: 20628129]
68. Giordano L, Leafman J, Citrin D, Wallace L. Cancer information-seeking practices among the Hispanic population: data from the Health Information National Trends Survey 2007. Hisp Health Care Int 2015 Jun;13(2):70-76. [doi: 10.1891/1540-4153.13.2.70] [Medline: 26078025]
69. Kreuter MW, Alcaraz KI, Pfeiffer D, Christopher K. Using dissemination research to identify optimal community settings for tailored breast cancer information kiosks. J Public Health Manag Pract 2008;14(2):160-169. [doi: 10.1097/01.PHH.0000311895.57831.02] [Medline: 18287923]
70. Allen JD, Mohllajee AP, Shelton RC, Drake BF, Mars DR. A computer-tailored intervention to promote informed decision making for prostate cancer screening among African American men. Am J Mens Health 2009 Dec;3(4):340-351 [FREE Full text] [doi: 10.1177/1557988308325460] [Medline: 19477736]
71. Eddens KS, Kreuter MW, Morgan JC, Beatty KE, Jasim SA, Garibay L, et al. Disparities by race and ethnicity in cancer survivor stories available on the web. J Med Internet Res 2009;11(4):e50 [FREE Full text] [doi: 10.2196/jmir.1163] [Medline: 19945948]
72. Greaney ML, Puleo E, Bennett GG, Haines J, Viswanath K, Gillman MW, et al. Factors associated with choice of web or print intervention materials in the healthy directions 2 study. Health Educ Behav 2014 Feb;41(1):52-62 [FREE Full text] [doi: 10.1177/1090198113486803] [Medline: 23720532]
73. Gustafson DH, McTavish FM, Stengle W, Ballard D, Hawkins R, Shaw BR, et al. Use and impact of eHealth system by low-income women with breast cancer. J Health Commun 2005 Jan;10(Suppl 1):195-218. [doi: 10.1080/10810730500263257] [Medline: 16377608]
74. Helft PR, Eckles RE, Johnson-Calley CS, Daugherty CK. Use of the internet to obtain cancer information among cancer patients at an urban county hospital. J Clin Oncol 2005 Aug 1;23(22):4954-4962 [FREE Full text] [doi: 10.1200/JCO.2005.09.621] [Medline: 16051948]
75. Kelly KM, Sturm AC, Kemp K, Holland J, Ferketich AK. How can we reach them? Information seeking and preferences for a cancer family history campaign in underserved communities. J Health Commun 2009 Sep;14(6):573-589. [doi: 10.1080/10810730903089580] [Medline: 19731128]
76. Kreuter MW, Black WJ, Friend L, Booker AC, Klump P, Bobra S, et al. Use of computer kiosks for breast cancer education in five community settings. Health Educ Behav 2006 Oct;33(5):625-642. [doi: 10.1177/1090198106290795] [Medline: 16923835]
77. Monnier J, Laken M, Carter CL. Patient and caregiver interest in internet-based cancer services. Cancer Pract 2002;10(6):305-310. [doi: 10.1046/j.1523-5394.2002.106005.x] [Medline: 12406053]
78. Paxton RJ, Nayak P, Taylor WC, Chang S, Courneya KS, Schover L, et al. African-American breast cancer survivors' preferences for various types of physical activity interventions: a Sisters Network Inc. web-based survey. J Cancer Surviv 2014 Mar;8(1):31-38 [FREE Full text] [doi: 10.1007/s11764-013-0307-5] [Medline: 24043292]
79. Quin J, Stams V, Phelps B, Boley T, Hazelrigg S. Interest in internet lung cancer support among rural cardiothoracic patients. J Surg Res 2010 May 1;160(1):35-39. [doi: 10.1016/j.jss.2008.10.004] [Medline: 19394639]
80. Ramirez TR, Mondragón D. Computer-based education for patients with cancer at Latino border hospital. Cancer Nurs 2002 Jun;25(3):245-250. [Medline: 12040234]
81. Santos SL, Tagai EK, Wang MQ, Scheirer MA, Slade JL, Holt CL. Feasibility of a web-based training system for peer community health advisors in cancer early detection among african americans. Am J Public Health 2014 Dec;104(12):2282-2289. [doi: 10.2105/AJPH.2014.302237] [Medline: 25320894]

82. Song H, Cramer EM, McRoy S. Information gathering and technology use among low-income minority men at risk for prostate cancer. Am J Mens Health 2015 May;9(3):235-246 [FREE Full text] [doi: 10.1177/1557988314539502] [Medline: 24951493]
83. Wise M, Han JY, Shaw B, McTavish F, Gustafson DH. Effects of using online narrative and didactic information on healthcare participation for breast cancer patients. Patient Educ Couns 2008 Mar;70(3):348-356 [FREE Full text] [doi: 10.1016/j.pec.2007.11.009] [Medline: 18201859]
84. Yi J, Zahn MP. Internet use among Chinese-American breast cancer survivors. Int Q Community Health Educ 2010 Jan;31(4):359-368. [doi: 10.2190/IQ.31.4.d] [Medline: 22192942]
85. McTavish FM, Gustafson DH, Owens BH, Hawkins RP, Pingree S, Wise M, et al. CHESS (Comprehensive Health Enhancement Support System): an interactive computer system for women with breast cancer piloted with an underserved population. J Ambul Care Manage 1995 Jul;18(3):35-41. [Medline: 10143478]
86. Johnson-Turbes A, Schlueter D, Moore AR, Buchanan ND, Fairley TL. Evaluation of a web-based program for African American young breast cancer survivors. Am J Prev Med 2015 Dec;49(6 Suppl 5):S543-S549 [FREE Full text] [doi: 10.1016/j.amepre.2015.09.003] [Medline: 26590650]
87. Chee W, Lee Y, Im E, Chee E, Tsai H, Nishigaki M, et al. A culturally tailored internet cancer support group for Asian American breast cancer survivors: a randomized controlled pilot intervention study. J Telemed Telecare 2016 Aug 2;23(6):618-626. [doi: 10.1177/1357633X16658369] [Medline: 27486198]
88. Kratzke C, Wilson S. Ethnic differences in breast cancer prevention information-seeking among rural women: will provider mobile messages work? J Cancer Educ 2014 Sep;29(3):541-547. [doi: 10.1007/s13187-013-0576-8] [Medline: 24163017]
89. Purnell JQ, Griffith J, Eddens KS, Kreuter MW. Mobile technology, cancer prevention, and health status among diverse, low-income adults. Am J Health Promot 2014;28(6):397-402 [FREE Full text] [doi: 10.4278/ajhp.120816-ARB-396] [Medline: 24200336]
90. Schoenberger YM, Phillips J, Mohiuddin MO, McNees P, Scarinci I. Acceptability of delivering and accessing health information through text messaging among community health advisors. JMIR Mhealth Uhealth 2013;1(2):e22 [FREE Full text] [doi: 10.2196/mhealth.2641] [Medline: 25099140]
91. Schoenberger YM, Phillips JM, Mohiuddin MO. Text messaging as a method for health ministry leaders to disseminate cancer information. J Cancer Educ 2015 Dec;30(4):636-641. [doi: 10.1007/s13187-014-0743-6] [Medline: 25355523]
92. Allen A, Hayes J. Patient satisfaction with teleoncology: a pilot study. Telemed J 1995;1(1):41-46. [doi: 10.1089/tmj.1.1995.1.41] [Medline: 10165321]
93. Doorenbos AZ, Eaton LH, Haozous E, Towle C, Revels L, Buchwald D. Satisfaction with telehealth for cancer support groups in rural American Indian and Alaska Native communities. Clin J Oncol Nurs 2010 Dec;14(6):765-770 [FREE Full text] [doi: 10.1188/10.CJON.765-770] [Medline: 21112853]
94. Hitt WC, Low G, Bird TM, Ott R. Telemedical cervical cancer screening to bridge medicaid service care gap for rural women. Telemed J E Health 2013 May 1;19(5):403-408 [FREE Full text] [doi: 10.1089/tmj.2012.0148] [Medline: 23600410]
95. Hitt WC, Low GM, Lynch CE, Gauss CH, Magann EF, Lowery CL, et al. Application of a telecolposcopy program in rural settings. Telemed J E Health 2016 Dec;22(10):816-820. [doi: 10.1089/tmj.2015.0260] [Medline: 27128600]
96. Mette LA, Saldívar AM, Poullard NE, Torres IC, Seth SG, Pollock BH, et al. Reaching high-risk underserved individuals for cancer genetic counseling by video-teleconferencing. J Community Support Oncol 2016 Apr;14(4):162-168. [doi: 10.12788/jcso.0247] [Medline: 27152515]
97. McDonald E, Lamb A, Grillo B, Lucas L, Miesfeldt S. Acceptability of telemedicine and other cancer genetic counseling models of service delivery in geographically remote settings. J Genet Couns 2014 Apr;23(2):221-228. [doi: 10.1007/s10897-013-9652-9] [Medline: 24014153]
98. Passik SD, Kirsh KL, Leibee S, Kaplan LS, Love C, Napier E, et al. A feasibility study of dignity psychotherapy delivered via telemedicine. Palliat Support Care 2004 Jun;2(2):149-155. [doi: 10.1017/S1478951504040209] [Medline: 16594244]
99. Pruthi S, Stange KJ, Malagrino GD, Chawla KS, LaRusso NF, Kaur JS. Successful implementation of a telemedicine-based counseling program for high-risk patients with breast cancer. Mayo Clin Proc 2013 Jan;88(1):68-73. [doi: 10.1016/j.mayocp.2012.10.015] [Medline: 23274020]
100. Bravo C, O'Donoghue C, Kaplan CP, Luce J, Ozanne E. Can mHealth improve risk assessment in underserved populations? Acceptability of a breast health questionnaire app in ethnically diverse, older, low-income women. J Health Dispar Res Pract 2014;7(4):6 [FREE Full text] [Medline: 25705576]
101. Arvey SR, Fernandez ME, LaRue DM, Bartholomew LK. When promotoras and technology meet: a qualitative analysis of promotoras' use of small media to increase cancer screening among South Texas Latinos. Health Educ Behav 2012 Jun;39(3):352-363 [FREE Full text] [doi: 10.1177/1090198111418110] [Medline: 21986243]
102. Banas JR, Victorson D, Gutierrez S, Cordero E, Guitleman J, Haas N. Developing a peer-to-peer mHealth application to connect Hispanic cancer patients. J Cancer Educ 2017 Dec;32(1):158-165. [doi: 10.1007/s13187-016-1066-6] [Medline: 27364905]
103. Smith SA, Whitehead MS, Sheats JQ, Fontenot B, Alema-Mensah E, Ansa B. Formative research to develop a lifestyle application (app) for African American breast cancer survivors. J Ga Public Health Assoc 2016;6(1):50-59 [FREE Full text] [doi: 10.21633/jgpha.6.103] [Medline: 27583307]

104. Weaver KE, Ellis SD, Denizard-Thompson N, Kronner D, Miller DP. Crafting appealing text messages to encourage colorectal cancer screening test completion: a qualitative study. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e100 [FREE Full text] [doi: 10.2196/mhealth.4651] [Medline: 26537553]
105. Le D, Holt CL, Saunders DR, Wang MQ, Coriolan A, Savoy AD, et al. Feasibility and acceptability of SMS text messaging in a prostate cancer educational intervention for African American men. Health Informatics J 2016 Dec;22(4):932-947. [doi: 10.1177/1460458215598636] [Medline: 26324051]
106. Berry DL, Halpenny B, Bosco JLF, Bruyere J, Sanda MG. Usability evaluation and adaptation of the e-health Personal Patient Profile-Prostate decision aid for Spanish-speaking Latino men. BMC Med Inform Decis Mak 2015 Jul;15:56 [FREE Full text] [doi: 10.1186/s12911-015-0180-4] [Medline: 26204920]
107. Im EO, Ji X, Zhang J, Kim S, Lee Y, Chee E, et al. Issues in developing and evaluating a culturally tailored internet cancer support group. Comput Inform Nurs 2016 Oct;34(10):462-469 [FREE Full text] [doi: 10.1097/CIN.0000000000000261] [Medline: 27379523]
108. Jaja C, Pares-Avila J, Wolpin S, Berry D. Usability evaluation of the interactive Personal Patient Profile-Prostate decision support system with African American men. J Natl Med Assoc 2010 Apr;102(4):290-297 [FREE Full text] [doi: 10.1016/S0027-9684(15)30601-5] [Medline: 20437736]
109. Joseph G, Beattie MS, Lee R, Braithwaite D, Wilcox C, Metrikin M, et al. Pre-counseling education for low literacy women at risk of Hereditary Breast and Ovarian Cancer (HBOC): patient experiences using the Cancer Risk Education Intervention Tool (CREdIT). J Genet Counsel 2010 May 19;19(5):447-462. [doi: 10.1007/s10897-010-9303-3] [Medline: 20490636]
110. May FP, Whitman CB, Varlyguina K, Bromley EG, Spiegel BM. Addressing low colorectal cancer screening in African Americans: using focus groups to inform the development of effective interventions. J Cancer Educ 2016 Sep;31(3):567-574 [FREE Full text] [doi: 10.1007/s13187-015-0842-z] [Medline: 25963898]
111. McTavish FM, Pingree S, Hawkins R, Gustafson D. Cultural differences in use of an electronic discussion group. J Health Psychol 2003 Jan;8(1):105-117. [doi: 10.1177/1359105303008001447] [Medline: 22113904]
112. Owens OL, Friedman DB, Brandt HM, Bernhardt JM, Hébert JR. An iterative process for developing and evaluating a computer-based prostate cancer decision aid for African American men. Health Promot Pract 2015 Sep;16(5):642-655 [FREE Full text] [doi: 10.1177/1524839915585737] [Medline: 25983084]
113. Kukafka R, Yi H, Xiao T, Thomas P, Aguirre A, Smalletz C, et al. Why breast cancer risk by the numbers is not enough: evaluation of a decision aid in multi-ethnic, low-numerate women. J Med Internet Res 2015 Jul 14;17(7):e165 [FREE Full text] [doi: 10.2196/jmir.4028] [Medline: 26175193]
114. Lee HY, Tran M, Jin SW, Bliss R, Yeazel M. Motivating underserved Vietnamese Americans to obtain colorectal cancer screening: evaluation of a culturally tailored DVD intervention. Asian Pac J Cancer Prev 2014;15(4):1791-1796 [FREE Full text] [doi: 10.7314/APJCP.2014.15.4.1791] [Medline: 24641410]
115. Menon U, Szalacha LA, Belue R, Rugen K, Martin KR, Kinney AY. Interactive, culturally sensitive education on colorectal cancer screening. Med Care 2008 Sep;46(9 Suppl 1):S44-S50 [FREE Full text] [doi: 10.1097/MLR.0b013e31818105a0] [Medline: 18725832]
116. Wang JH, Liang W, Schwartz MD, Lee MM, Kreling B, Mandelblatt JS. Development and evaluation of a culturally tailored educational video: changing breast cancer-related behaviors in Chinese women. Health Educ Behav 2008 Dec;35(6):806-820. [doi: 10.1177/1090198106296768] [Medline: 17602099]
117. Owens OL, Friedman DB, Brandt HM, Bernhardt JM, Hébert JR. Digital solutions for informed decision making: an academic-community partnership for the development of a prostate cancer decision aid for African American men. Am J Mens Health 2016 Dec;10(3):207-219 [FREE Full text] [doi: 10.1177/1557988314564178] [Medline: 25563381]
118. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
119. Pew Research Center. 2013. Who's not online and why URL: http://www.pewinternet.org/2013/09/25/whos-not-online-and-why/ [accessed 2019-02-26] [WebCite Cache ID 76UUlbp0g]
120. Chou WS, Liu B, Post S, Hesse B. Health-related internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
121. Pew Research Center. 2015. US Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2019-02-26] [WebCite Cache ID 76UDFCeC7]
122. Xie B, Bugg JM. Public library computer training for older adults to access high-quality internet health information. Libr Inf Sci Res 2009 Sep 1;31(3):155-162 [FREE Full text] [doi: 10.1016/j.lisr.2009.03.004] [Medline: 20161649]
123. Xie B. Improving older adults' e-health literacy through computer training using NIH online resources. Libr Inf Sci Res 2012 Jan 1;34(1):63-71 [FREE Full text] [doi: 10.1016/j.lisr.2011.07.006] [Medline: 22639488]
124. Street RL, Gold WR, Manning TR. Health Promotion and Interactive Technology: Theoretical Applications and Future Directions. Hoboken: Taylor and Francis; 2013.
125. Kreuter MW, Farrell DW, Olevitch LR, Brennan LK. Tailoring Health Messages: Customizing Communication with Computer Technology. Hoboken: Taylor and Francis; 2013.

126. Betancourt J, Green A, Carrillo J, Ananeh-Firempong O. Defining cultural competence: a practical framework for addressing racial/ethnic disparities in health and health care. Public Health Rep 2003;118(4):293-302 [FREE Full text] [doi: 10.1093/phr/118.4.293] [Medline: 12815076]
127. Chang BL, Bakken S, Brown SS, Houston TK, Kreps GL, Kukafka R, et al. Bridging the digital divide: reaching vulnerable populations. J Am Med Inform Assoc 2004;11(6):448-457 [FREE Full text] [doi: 10.1197/jamia.M1535] [Medline: 15299002]
128. Luger TM, Hogan TP, Richardson LM, Cioffari-Bailiff L, Harvey K, Houston TK. Older Veteran digital disparities: examining the potential for solutions within social networks. J Med Internet Res 2016 Nov 23;18(11):e296 [FREE Full text] [doi: 10.2196/jmir.6385] [Medline: 27881361]
129. Siegel R, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018 Dec;68(1):7-30 [FREE Full text] [doi: 10.3322/caac.21442] [Medline: 29313949]
130. DeSantis C, Siegel R, Sauer A, Miller K, Fedewa S, Alcaraz K, et al. Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. CA Cancer J Clin 2016 Dec;66(4):290-308 [FREE Full text] [doi: 10.3322/caac.21340] [Medline: 26910411]
131. Siegel R, Fedewa S, Miller K, Goding-Sauer A, Pinheiro P, Martinez-Tyson D, et al. Cancer statistics for Hispanics/Latinos, 2015. CA Cancer J Clin 2015;65(6):457-480 [FREE Full text] [doi: 10.3322/caac.21314] [Medline: 26375877]
132. Connecting for Health. Personal Health Records and Sharing Patient Information. New York: Markle Foundation; 2006.
133. Healthcare Information Management and Systems Society. 2007. HIMSS personal health records definition and position statement URL: http://www.himss.org/content/files/PHRDefinition071707.pdf [accessed 2019-02-26] [WebCite Cache ID 76UHXyZkH]
134. King JS, Eckman MH, Moulton BW. The potential of shared decision making to reduce health disparities. J Law Med Ethics 2011 Mar;39(Suppl 1):30-33. [doi: 10.1111/j.1748-720X.2011.00561.x] [Medline: 21309892]
135. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]

Abbreviations

eHealth: electronic health
HIT: health information technology
mHealth: mobile health
PHR: personal health record
RCT: randomized controlled trial
SDM: shared decision making
SMS: short message service
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 28.02.18; peer-reviewed by YM Schoenberger, X Cheng, D Du; comments to author 12.04.18; revised version received 31.08.18; accepted 14.09.18; published 23.04.19.

Please cite as:
Tarver WL, Haggstrom DA
The Use of Cancer-Specific Patient-Centered Technologies Among Underserved Populations in the United States: Systematic Review
J Med Internet Res 2019;21(4):e10256
URL: http://www.jmir.org/2019/4/e10256/
doi:10.2196/10256
PMID:31012855

Original Paper

Patients' Needs and Requirements for eHealth Pain Management Interventions: Qualitative Study

Ingrid Konstanse Ledel Solem[1,2], MS; Cecilie Varsi[1], PhD; Hilde Eide[1,3], PhD; Olöf Birna Kristjansdottir[1,4], PhD; Jelena Mirkovic[1], PhD; Elin Børøsund[1], PhD; Mette Haaland-Øverby[1,4], BS; Karina Heldal[1,2], MS; Karlein MG Schreurs[5], PhD; Lori B Waxenberg[6], PhD; Karen Elizabeth Weiss[7], PhD; Eleshia J Morrison[8], PhD; Lise Solberg Nes[1,2,9], PhD

[1]Center for Shared Decision Making and Collaborative Care Research, Division of Medicine, Oslo University Hospital, Oslo, Norway
[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[3]Science Centre Health and Technology, University of South-Eastern Norway, Drammen, Norway
[4]Norwegian National Advisory Unit on Learning and Mastery in Health, Oslo University Hospital, Oslo, Norway
[5]Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands
[6]Department of Clinical and Health Psychology, University of Florida, Florida, FL, United States
[7]Department of Anesthesiology and Pain Medicine, University of Washington School of Medicine, Washington, WA, United States
[8]Mayo Clinic, Pain Rehabilitation Center, Rochester, MN, United States
[9]Department of Psychiatry & Psychology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Lise Solberg Nes, PhD
Center for Shared Decision Making and Collaborative Care Research
Division of Medicine
Oslo University Hospital
Oslo University Hospital HF, Division of Medicine, Aker hospital
Box 4950 Nydalen
Oslo, N-0424
Norway
Phone: 47 23 06 60 04
Email: lise.solberg.nes@rr-research.no

Abstract

Background: A growing body of evidence supports the potential effectiveness of electronic health (eHealth) interventions in managing chronic pain. However, research on the needs and preferences of patients with chronic pain in relation to eHealth interventions is scarce. Eliciting user input in the development of eHealth interventions may be a crucial step toward developing meaningful interventions for patients for potentially improving treatment outcomes.

Objective: This study aimed to explore the experiences of patients with chronic pain with regard to information and communication technology, understand how an eHealth intervention can support the everyday needs and challenges of patients with chronic pain, and identify possible facilitators and barriers for patients' use of an eHealth pain management intervention.

Methods: Twenty patients living with chronic pain and five spouses participated in individual interviews. Semistructured interview guides were used to explore participants' needs, experiences, and challenges in daily life as well as their information and communication technology experiences and preferences for eHealth support interventions. Spouses were recruited and interviewed to gain additional insight into the patients' needs. The study used qualitative thematic analysis.

Results: The participants were generally experienced technology users and reported using apps regularly. They were mainly in favor of using an eHealth self-management intervention for chronic pain and considered it a potentially acceptable way of gathering knowledge and support for pain management. The participants expressed the need for obtaining more information and knowledge, establishing a better balance in everyday life, and receiving support for improving communication and social participation. They provided suggestions for the eHealth intervention content and functionality to address these needs. Accessibility, personalization, and usability were emphasized as important elements for an eHealth support tool. The participants described an ideal eHealth intervention as one that could be used for support and distraction from pain, at any time or in any situation, regardless of varying pain intensity and concentration capacity.

Conclusions: This study provides insight into user preferences for eHealth interventions aiming to address self-management for chronic pain. Participants highlighted important factors to be considered when designing and developing eHealth interventions for self-management of chronic pain, illustrating the importance and benefit of including users in the development of eHealth interventions.

Trial Registration: ClinicalTrials.gov NCT03705104; https://clinicaltrials.gov/ct2/show/NCT03705104.

(J Med Internet Res 2019;21(4):e13205) doi:10.2196/13205

KEYWORDS

chronic pain; eHealth; self-management; qualitative methods; intervention development; user-centered design

Introduction

Chronic pain conditions are common (affecting 25%-30% of the population) and difficult to cure, with a significant impact on the persons affected and on society in terms of economical, psychological, and social issues [1-3]. Like all long-term conditions, chronic pain requires day-to-day self-management by those affected. This includes managing the emotional and medical consequences of the chronic condition; self-regulatory efforts; and changing, maintaining, and creating new sets of behaviors to improve coping mechanisms [3-5]. Supporting patients and caregivers in the self-management process is an important step toward better health care services and have shown to have a positive impact on patient outcomes, including decreased pain interference and pain intensity and improved self-efficacy [6-9].

A growing body of evidence supports the potential efficacy of electronic health (eHealth) interventions contributing to self-management of chronic pain [10-16], which refer to interventions using information and communication technology (ICT) such as apps, websites, or remotely delivered interventions/telehealth or telecare in the delivery of health care services [17]. eHealth interventions have the potential to make health care services more available to patients, allowing patients to access services and help from their own home. In addition, eHealth interventions may introduce more cost-effective treatment options, reducing the need for travel and direct health care personnel involvement [13,18]. Such interventions also have the potential to enhance treatment durability, as patients can receive support and reinforcement of skills during and after treatment [13].

Patients with chronic pain have also shown interest in eHealth interventions [19]. For some, such interventions could even be the preferred option, as they are easily accessible, possibly perceived as neutral and nonjudgmental, and allow patients to continue treatment and support at their own pace [20]. There is, however, a gap between the commercial and scientific aspects of eHealth tools. Available apps for people with chronic pain commonly focus on physical health and include a functionality supporting monitoring/tracking, assessment, feedback, and information/education [21]. Surprisingly, a few existing apps appear to be based on theoretical and evidence-based rationale, and a few appear to be developed or evaluated using scientific methods [21-23]. In addition, few eHealth interventions are developed by, or in collaboration with, health care professionals [12,23,24].

The success of eHealth interventions depends on technology and content but, perhaps, just as much on patients’ acceptance of and adherence to the intervention [25]. Involving patients in the process appears essential in the development of effective eHealth interventions. However, several studies have pointed out the lack of user involvement in the development of such interventions [23,25-27]. Keogh and colleagues [26] described eHealth as a promising area for pain management but emphasized the need to maintain patient focus and recommended using user-centered designs as the starting point, involving patients (called users) in the entire development process [26]. The authors also stated that eHealth interventions, even when claiming to be therapeutic, are often developed in response to a technological innovation, rather than user needs [26]. Other studies have pointed out that including users in the design of eHealth interventions would allow for tailoring of individual preferences [28], and personalization and tailoring in such interventions may improve impact [20,23]. Despite these recommendations from existing research, users are rarely involved early on in eHealth development processes [25]. A recent study addressed this issue by combining a review of the literature with a focus group study including patients with chronic pain, their caregivers, and their health care providers. This resulted in a suggestion of elements needed for an eHealth intervention for people with chronic pain, including enriched information environment, automated tailored feedback observations of individual progress, automated follow-up messages, communicative function (eg, advisor or peer-support access), and use of supplementary modes (eg, access to and use of a variety of materials) [29].

The aims of this study were (1) to advance knowledge regarding the needs and requirements for eHealth pain management interventions to explore experiences of patients with chronic pain with regard to information and communication technology (ICT) in order to explore how an eHealth intervention can support the everyday needs and challenges of patients with chronic pain and (2) to identify possible facilitators and barriers for patients’ use of an eHealth pain management intervention. This study is the first step in a larger project where the aim is to design, develop, and test a user-centered eHealth intervention for adults with chronic pain based on cognitive behavioral therapy (CBT; trial registration: NCT03705104).

Methods

Study Design

This study used a qualitative design involving individual interviews with patients with chronic pain and their spouses [30], to explore patients' needs and preferences for designing and developing eHealth interventions. In the study, spouses were included to gain additional insight into the perceived needs of patients, from the spouses' perspective. Identifying and exploring patients' needs, experiences (in everyday life as well as in relation to technology), and preferences are early steps in a user-centered development process [31,32].

Recruitment

To be eligible for study participation, participants had to be 18 years or older, have experienced chronic pain for 3 months or more, and be able to communicate in Norwegian. Inclusion criteria for spouses were that they were married or cohabitating with one of the participating patients. Recruitment was conducted by collaborative health care providers at four collaborative institutions. Persons who met the inclusion criteria were invited to participate in individual interviews by health care providers at local patient education centers, pain clinics, physical therapy institutes, and psychology practices. Potential participants first received information about the study through hand-out pamphlets and from their health care providers. If the potential participants were interested, they were contacted by a researcher for more information. Of the people who had agreed to be contacted by the research team, none declined participation. Spouses were contacted after obtaining consent from the patients, only if the participating patient agreed that their spouse could be contacted. Participants (patients and spouses) were offered a gift certificate (value of approximately US $30) as compensation for their time.

Ethical Approval and Informed Consent

The study was approved by the Institutional Review Board (approval number: 2017/6697) at a major medical center in northern Europe. Informed consent was obtained after the participants were given information about the nature of the study and aspects of participation.

Data Collection

Although this study was exploratory and open in nature, the research was conducted with a specific outcome—the development of an eHealth intervention. The interview and analyses were therefore guided by the aims of the study. Semistructured interview guides (Multimedia Appendix 1) were developed by the research team and used to explore participants' (patients' *and* spouses') everyday life with chronic pain, including everyday routines, challenges, and coping strategies; participants' experiences and thoughts about technology and health-related technology, including smartwatches, activity and nutrition trackers, mindfulness apps/videos, health forums and blogs, and other health-related apps; and participants' thoughts and expressed needs related to an eHealth pain management intervention. Participants (ie, patients and spouses) were interviewed individually to capture patients' and spouses' views separately. Participants were also asked to complete a brief questionnaire assessing background information, including age, type of diagnosis, and treatments in addition to experiences with technology. A patient representative provided input on the background questionnaire and the interview guides to ensure easily understandable and nonoffensive questions. The interviews were audio recorded and conducted face to face by the first author (ILS) and conducted either at the research center, in meeting rooms at local patient education centers, in participants' home, or at participants' workplaces. The first two interviews were conducted by the first author together with an experienced interviewer (CV).

Data Analysis

Interview data were transcribed verbatim and analyzed using a thematic analysis approach in a stepwise process [33]. In the first step, data were deductively grouped into broad themes derived from CBT themes (eg, health-promoting behaviors, thoughts and feelings, and social relations). Second, sentences and longer semantic units were coded and grouped into categories across and within the initial themes. This was done on a manifest as well as a latent level, so that what the respondents actually said as well as the underlying assumptions were analyzed [33]. Each category was then examined again, with a focus on identifying variations, similarities, and differences within each category. Subcategories (ie, minor themes) were identified and named by characterizing content. Finally, all subcategories were examined again and grouped into main categories (ie, major themes). To ensure trustworthiness, the research team consisting of the first author (ILS) and coauthors (LSN, CV, OK, and HE) met regularly to discuss and refine the analysis until an agreement was reached [34]. NVivo 11 (QSR International, Victoria, Australia) qualitative analysis software was used to organize and facilitate the analysis. To increase the transparency of the interpretation, categories and subcategories were illustrated with quotations.

Results

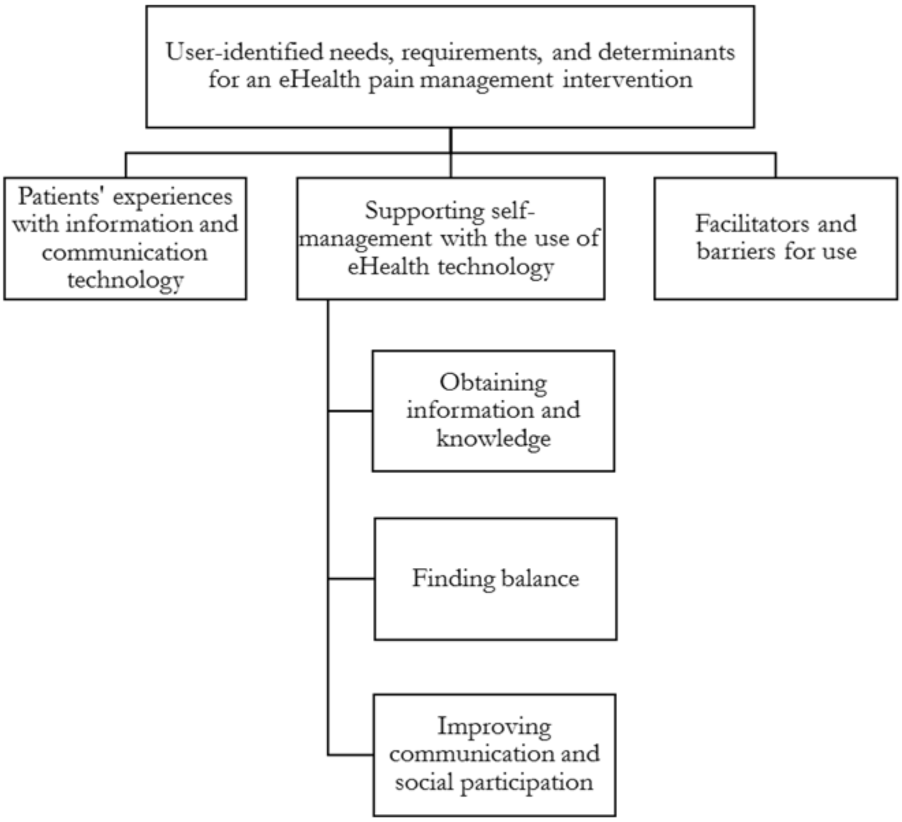
Overview

A total of 20 persons (15 women) living with different types of pain and 5 spouses (2 women) were included in the study. The participants' (patients' *and* spouses') age ranged from 18-74 years, with a median of 48 years. Many patients noted that it took some years before they received a diagnosis, with some having lived with pain for 10 years or more before reportedly getting their diagnosis. Almost all patients had been through a variety of treatments, ranging from primary care (eg, general practitioners' visits and physical therapy) to more specialized treatments and rehabilitation in secondary and tertiary care. There were no clear differences in the results between male and female participants in this study, apart from women describing their pain by using more metaphors. Table 1 presents a detailed view of the patient demographics.

Table 1. Patient demographics (N=20).

Characteristic	Number of patients (%)
Employment status	
Working/studying full-time	5 (25)
Working/studying part-time	4 (20)
Currently on sick leave	5 (25)
On disability benefits	6 (30)
Type of pain	
Neck and back pain	8 (40)
Neurological pain	8 (40)
Others	4 (20)
Reported time living with pain (years)	
0-3	2 (10)
4-8	5 (25)
9-15	5 (25)
16-25	4 (20)
≥26 years	4 (20)

Figure 1. Overview of the main categories and subcategories from the analysis. ICT: information and communication technology.



The findings provided (1) insight on patients’ experiences with ICT, (2) understanding of how an eHealth intervention could support patients’ everyday needs and challenges (ie, support self-management), and (3) information on the facilitators and barriers for patients’ use of an eHealth pain management intervention. Categories 1 and 3 were analyzed on a manifest (visual) level, while category 2 was analyzed on a latent (interpretative) level. Main categories and subcategories are illustrated in Figure 1.

Patients' Experiences with Information and Communication Technology

All participants reported owning a smartphone, computer, or tablet and stated that they were using apps regularly, mostly for practical day-to-day use such as buying a bus ticket or for checking the weather forecast. Nearly everyone had previously downloaded an app, and about half (n=11) reported having used a health-related app, mainly mindfulness apps for relaxation and focus control, or apps for step counting. Those who had tried apps for step counting often stopped using them, as they felt as if they were never able to walk enough because of their pain. Some, especially those under 35 years of age, also regularly used apps for podcasts as a strategy to distract focus away from their pain. Only two participants had tried apps specifically for pain management, mainly with a focus on migraine, and rarely, apps were offered in the participants' native language. Many stated that they did not find or know of any apps specifically for pain management. A few participants had used YouTube for videos showing mindfulness, breathing, and visualization exercises. Mindfulness exercises (using apps, YouTube, or CDs) were particularly used at night when in bed as a relaxation strategy when trying to sleep. Participants who had not tried any health-related programs (eg, apps, smart watches, or similar) reported not having thought about it, not finding any apps relevant enough for their needs, or not knowing what to download and use.

The majority of the participants described using their mobile phone more than their tablets and computers. They preferred to use their phone for apps, as they found that apps were more easily available on the phone. Tablets were used when searching for information or reading news, while computers were mostly used at work or for practical tasks such as paying bills. Some also reported using their computer for watching television series, a strategy often used for distraction from pain. Searching for information about pain and treatments on the internet was common, especially during the first years of living with chronic pain, but many emphasized challenges related to finding information on the internet, as they did not know what information to trust, and expressed concern about receiving incorrect or even damaging information or advice. Most participants had checked out different pain forums, blogs, or groups on the internet, but few reported using or participating in these actively; instead, the patients perceiving them as "too negative" and "depressing" to follow . Some (n=5) found these types of forums useful when needing advice from peers, emphasizing that they avoided looking at these forums on days with a lot of pain due to the "negative vibes" from these forums. Only three participants reported communicating with their health care providers electronically (ie, through email or patient/physician web pages).

Participants were generally in favor of using an eHealth self-management intervention for chronic pain: Only one patient did not visualize how or when such an intervention could be of use. Participants were enthusiastic about potentially getting an easily accessible tool in their daily lives that could be used for support and distraction from pain, at any time or in any situation. Thus, a mobile app was identified by the participants early on as the preferred platform for use.

Supporting Self-Management With the Use of eHealth Technology

Overview

Analyses showed that participants expressed the need for self-management support. Independent of the type of pain, participants (both patients and spouses) described common challenges in their daily lives including physical, psychological, and social challenges such as fatigue, isolation, depressive thoughts and anxiety, sorrow and guilt, and memory and concentration problems. The expressed needs, including those regarding an eHealth intervention, appeared related to years lived with chronic pain and their "good or bad days," which depended on daily pain intensity and level of fatigue.

Taking responsibility and experiencing independence were viewed by participants as essential for self-worth, but challenging without the necessary information, knowledge, and support. The informants' expressions for self-management support through the use of eHealth technology were related to three different aspects: obtaining information and knowledge, finding balance, and receiving support for improving communication and social participation.

Obtaining Information and Knowledge

Most participants (ie, patients and spouses) were positive toward, or expressed a wish for, receiving information and gaining knowledge through technology and educational texts, communication with health care professionals, or both. Topics identified included the need for information about pain, in general, and pain physiology; health-promoting behaviors including information about sleep, activity pacing, physical activity, and nutrition; treatment options, medication, and medical aids; psychosocial information including information about communicating pain with others and coping with anxiety and depressive thoughts; and reports on the newest pain-related research.

Many of the participants reported that a lack information about pain and pain management, especially during their first years living with pain, induced a feeling of desperation and anxiety. For instance, many patients had worried about their memory issues, with catastrophizing thoughts of having dementia. As a result, many stated that they had been "shopping" for health care services and alternative treatments, trying "everything" they came across instead of focusing on pain management strategies for themselves. Over the years, many had found their own ways of managing their pain, stating that they did not have the same need for information now as compared to early in the disease trajectory. These patients further emphasized that they wished they had received this kind of information earlier, instead of trying to work out all these things on their own. Compared to the patients who had not lived with pain for very long, these patients wanted more information on the newest research, knowledge, coping strategies, and exercises that could support positive psychological well-being.

Many participants expressed a feeling that health care providers, in meetings with them, lacked knowledge about their pain or gave conflicting information about what to do with the pain. Several participants were told by their physician that chronic pain is difficult to deal with and some were instructed to look up chronic pain on the internet themselves, resulting in further contradictory and sometimes scary information. Two participants had initiated additional education for themselves (ie, studying psychotherapy and medicine) in an attempt to gain an in-depth understanding of their pain and how to best cope with it.

The participants also emphasized the need for independence. Several preferred to work things out by themselves, if possible, yet needed more knowledge to do so successfully. Getting trustworthy information from health care providers and researchers in a self-management app was seen as a positive option that could potentially provide the needed knowledge. Some wanted direct contact with, and support from, health care providers through the intervention, but expressed doubt about how this would work, considering the need for contact with specialized health care providers who were already extremely busy and often unavailable. Participants also described the daytime work hours of health care personnel as challenging, as patients often need help at different hours of the day, not just during daytime hours. However, both patients and spouses considered the potential of having access to information and knowledge through an eHealth intervention as beneficial. The spouses pointed to personal challenges in understanding the pain and suffering caused by pain. They wanted more information about the pain and how it could connect to their partner's psychological health and well-being, emphasizing the need for knowledge about how to better communicate with each other despite strong emotions or pain. They also described feelings of being left out and helplessness seeing their partner in pain and not being able to adequately help.

The potential for an eHealth intervention to provide information on the patient's own terms was further emphasized, as the ability to decide when and how much information to receive was considered focal. Table 2 provides a summary of participants' expressed needs and suggestions for content and functionality, and Multimedia Appendix 2 provides additional quotes from the participants.

Finding Balance

A topic addressed by all participants (ie, patients and spouses) was the need for, and the challenges around, finding a balance in everyday life. This particularly included challenges with finding a balance between seeking help versus being independent, being active while getting enough rest (ie, activity pacing), and focusing on and talking about the pain versus distraction from the pain. The majority of the participants had accepted their pain and did not think that it would ever subside, although they found it difficult to cope with and accept the *consequences and impact* of pain on their daily lives. They described challenges around fully understanding how pain is affected by other variables in life, such as sleep, mood, and activity. Despite being familiar with the concept of activity pacing, few participants actually practiced such pacing, stating a lack of competence and knowledge of how to appropriately pace activities. Those who mastered activity pacing had achieved this mastery through repeatedly trying and failing over many years of living with pain, and they wished they had learned about the importance of activity pacing early on. The patients also described challenges about making others understand their pain. They wanted support and advice in this process, so that they could become more balanced and better cope on a day-to-day basis in their pursuit of regaining a "normal" or "new normal" life.

> *It's important to feel normal and it's important to have a NORMAL everyday life. The daily life from before is not yours anymore. You need to find a NEW normal. And in that process, it is important to have a tool, important to have...overview.* [patient]

Some of the patients had positive experiences with relaxation exercises, stating that they were helping them feel "more balanced." They wanted such exercises made more available in everyday life, for instance, through a mobile app, specifying that reminders would be helpful, as they often forgot to do such exercises. Others wanted help with focusing on the positive things in life, suggesting that some forms of "words of wisdom" might be useful.

One topic addressed by many was the possibility of doing daily registrations. The patients wanted to be able to register details such as mood, sleep, pain, and activity in order to gain better control over their own coping strategies and skills, rather than always relying on health care services. Providing patients with visual information, for instance, in a graph consisting of their registrations, could give them concrete information that they could learn from, but also share with their partner or health care provider, making it easier for those around them to understand how the pain affects them.

Some of the patients, and especially the spouses, also emphasized that such registrations could be a way of putting things into perspective, as registrations could show them that a variable such as sleep or a situation such as a social event, might not have the impact they thought it would have. One spouse said:

> *During especially bad times, one feels that this is worse than ever before. And that this will never be better or OK again. And if you then had some clear statistics, something objective. Something that could say; "you were actually doing worse this time last year. But just a month later, you climbed that mountain".* [spouse]

Some spouses emphasized that their partner did more than they thought they did, often pushing themselves too hard, sometimes even harder than people without health issues. Registrations could help patients be more satisfied with a lower or a more balanced activity level, suggesting that registrations could support acceptance in daily life. At the same time, some participants also recognized that such registrations could have a negative impact on their mood and self-efficacy (eg, not experiencing "good days" and hence only registering "bad days").

Table 2. Participants' (N=25) needs, suggestions, and reasoning for content and functionality of eHealth interventions.

Needs	Suggestions for content/functionality	Reasoning
Information and knowledge	Information about pain physiology, treatments and medicine, and legal rights and pain management	• Feel safe • Feel less uncertainty and desperation • Be more independent • Help accept the consequences of the pain
	Direct contact with health care providers	• Get to ask questions when needed • Feel safer • Get constructive feedback for better coping and motivation
Balance	Daily registrations of variables like sleep, mood activity, and pain	• See how the variables connect and learn from mistakes • Get a broader perspective • Something to show health care providers/partners for better support
	Pain diary/notes	• Save important messages, advice, and experiences from doctor's appointments, courses, etc • Get a broader perspective on personal thoughts and experiences and get help accepting the consequences of the pain
	Calendar	• Keep track of appointments regarding pain treatments, to avoid your personal calendar getting too pain focused
	Medical diary/list	• Keep track of medical history • Bring to health care providers
	Word of the day/word of wisdom	• A reminder of what's most important in life • Motivation
Communication and social participation	Social forum or inspirational stories from peers	• Feel less alone • Get support and advice from peers • Become more motivated for change
	Writing three positive things/self-praise	• Better mood/be more enjoyable to be with • A reminder of what's most important in life • Focus on the things one is able to manage
	Advice on communication or direct contact with partner via the app	• Be able to talk about pain at home, possibly experiencing more understanding and support
	Breathing, focus, and relaxation exercises	• As a break/distraction from the pain • To help become more present in daily activities and social gatherings

Some patients had tried such registrations during clinical pain assessments and treatments or in relation to applying for disability benefits, but found it challenging to remember all these variables or carry a notebook all the time. Using the phone for such registrations was therefore considered an easier and more accessible option. Table 2 provides more details on this, and Multimedia Appendix 2 provides additional quotes from the participants.

Improving Communication and Social Participation

Patients' needs for self-management included a need for normalization and independence. This also influenced their social relations, and the majority described challenges with accepting role changes and finding their new place in social settings. Many described pushing themselves in the attempt to be "like they used to be."

> *If I'm not given special attention [due to the pain], I've actually achieved what I'm dreaming of. Then I have managed to stabilize everything, so that life is normal. But it's only a dream.* [patient]
>
> *I'm the mom, NOT the sick mother. That's extremely important to me...I do everything to be able to be and say that.* [patient]

Many participants (spouses and patients) described that these attempts of trying to be "like they used to be," often resulted in increased pain and fatigue, with the patients becoming more distant, irritated, and sometimes angry. The spouses found this difficult to deal with and emphasized that their partner could use help with acknowledging the situation, so that they would not push themselves so hard or so that they could open up to the idea of medical aids and supporting tools. The patients, on the other hand, expressed difficulties getting their partner to understand their needs for normality and independence. One patient emphasized that it was “all about how things were said.” She had experienced positive changes at home after her husband had participated in a self-management course for caregivers. She described him as having gone from being frustrated and accusatory toward her whenever she did too much, to clearly expressing his *own wish* to do those things, thereby making it easier for her to let go. This particular participant, as well as others, emphasized the need for openness, even though most of the participants stated that they did not talk much about their pain at home. They stressed upon the need for a conversation starter and for advice on how to make necessary adjustments to facilitate such conversations.

High pain levels, concentration issues, and fatigue were considered to hinder patients’ social participation, as many felt like they were “not present enough” or that they were “too aggressive.” The need for something that could distract and give them a break from their pain was emphasized. A reminder to perform a breathing exercise, perhaps, after work before sitting down for dinner with the family, was one suggestion. Several stated that they did not want health care providers to follow their progress in these tasks, as they felt this could pressurize them in a negative way, since they already feel as if they never did enough.

Some participants also suggested that being able to connect with peers through a type of social forum could be a useful functionality, making them feel less alone and less frustrated. They did, however, state that they did not want any form of “competition” with peers. Several hoped for inspiration from others through social forums and to learn from their experiences, either directly through a forum or as stories to read and choose from. When asked why they imagined using a forum in the eHealth intervention when they did not want to participate in forums/pain groups online, some answered that they thought this could be a more positive experience, as the intervention likely would target those who wanted to learn to cope with the pain, rather than “complain about it.” Table 2 provides more details on this, and Multimedia Appendix 2 provides additional quotes from the participants.

Facilitators and Barriers for Use

Participants were also asked to reflect upon what might make an app “good or bad” in their opinion and any thoughts they might have about potential facilitators and barriers for use of an eHealth self-management intervention for chronic pain. A summary of topics discussed and emphasized by the participants around this are presented in Table 3 (for additional quotes from the participants, see Multimedia Appendix 2). Participants seemed to agree that “good apps” are those that are user-friendly and easy to use; informative and functional, without being overwhelming (several participants had stopped using apps because they felt they were too overwhelming); and without errors or too many updates. Regarding an eHealth self-management intervention, accessibility was addressed as an important facilitator to allow patients to easily access the coping resources they needed when experiencing elevated pain and fatigue levels. Suggestions included having a simple login procedure, emphasized by many as a necessity for use. Between options of a variety of functionalities (eg, a journal, forum, and social connection) or a simple login, 15-20 patients chose a simple login.

Participants did, however, emphasize the importance of having a variety of options to choose from, related to design features and content units, particularly considering their pain and concentration challenges, making the intervention more personalized. For example, they suggested that the intervention could give them advice about content, topic, and exercises to perform based upon their daily registrations. Alternatively, they wanted to be able to choose the content topic and duration of use, for instance, through the use of a “read more” button and an exercise list. This could also allow them to choose exercises or brief educational texts on “bad days” with high pain and concentration issues. The participants also emphasized the need for reliable and up-to-date information and stated that they had chosen to participate in this study, as they felt safe knowing that health care professionals were involved in the intervention development. Poor usability, including nonintuitive design and an overwhelming amount of information, was emphasized as the main barriers for use.

Table 3. Facilitators and barriers for the use of eHealth interventions.

Themes/topics	Facilitators	Possible barriers
Accessibility and privacy	Mobile app for everyday use, tablet for longer reads	Computer and tablet not accessible enough throughout the day
	Simple login; code is acceptable and trustworthy	Cumbersome login procedure
Usability	Simple design with intuitive icons	Complex icons or background noise that could disturb the concentration level
	Easily readable and short texts, preferably with a "read more" button	Longer texts that challenge the concentration level
Personalization and tailoring	Daily registrations (ie, sleep, mood, and pain) for more personalized content based on needs and challenges	Overall lack of personalization
	Reminders: ability to choose when and how	Continuous reminders with "bad timing"
	Voice-overs: possibility of choosing between different voices	A "wrong" voice
Reliability of the intervention	Evidence-based content and involvement of specialized health care professionals	None
	Updates and up-to-date information	Never any new content
	Supportive and nonjudgmental language	Judgmental, negative, or "glossy" language

Discussion

Principal Findings

Identifying and exploring patients' needs, experiences, and preferences are of essence when designing and developing eHealth interventions. In this study, patients with chronic pain and some of their spouses participated in the early stages of developing an eHealth pain management intervention. The participants (ie, patients and spouses) provided insight into patients' experiences (and from the spouses' point of view, patients' perceived experiences) with ICTs, their needs and challenges in relation to an eHealth intervention, and their thoughts about possible facilitators and barriers for the use of an eHealth pain management intervention.

Participants were generally in favor of using an eHealth self-management intervention for chronic pain and considered such a potential eHealth tool acceptable for gathering knowledge and gaining support related to pain management. Despite previous experience with health-related apps (eg, mindfulness apps and exercising apps) or online pain groups, participants described a lack of existing eHealth apps and interventions specifically targeting chronic pain and pain management. Participants were enthusiastic about the prospect of obtaining a tool targeting pain management and receiving information as well as useful exercises, and through them, "getting everything you need in one place." Easy access and availability were regarded necessities for use, with participants depicting a mobile app as the preferred platform for use.

Regardless of the type of pain, participants described comparable challenges in their daily lives, including fatigue and sleeping challenges; memory and concentration issues; and psychosocial challenges such as negative thoughts, anxiety, guilt, sorrow, and feelings of isolation, all supporting the existing literature [1,2,35]. Participants expressed a need for obtaining more information and knowledge, finding a balance in everyday life, and receiving support for improving communication and social participation. For an intervention to best address and facilitate support for these needs, the participants also provided suggestions for the eHealth intervention content and functionality.

Possible facilitators and barriers for patients' use of an eHealth intervention were also identified, with participants emphasizing that an accessible tool could be used in any circumstance, regardless of the varying pain level or concentration issues, as a likely facilitator for use. Poor usability and limited or lacking tailoring or personalization were identified as the potential main barriers for use.

Providing Self-Management Support Through eHealth

Acceptance is considered a necessary component when teaching self-management skills, as individuals who have accepted their pain are more open and willing to take an active role in the self-management process [7]. The majority of the patients in this study described having accepted their pain, suggesting that they were open for self-management education and support. This was also reflected in their interest in eHealth interventions for self-management and their expressed need for self-management strategies. The participants (ie, patients and spouses) expressed needs that reflect common content and themes from existing face-to-face CBT interventions targeting chronic pain, including a wish for more knowledge about recommended health-related behaviors, activity pacing, and psychosocial support. Patients also emphasized a need for independence and normalization, yet expressed a wish for support, or as one participant said, "some tools" that could be helpful in this process. This supports the notion that patients prefer self-management support [4] in order to respond to physical and mental changes and to manage their day-to-day challenges and decisions.

Patient empowerment involves a process to enable patients to have more influence over their health by promoting their capacities to gain control over self-defined important matters, thus leading to better self-management [36]. As such, supporting patients in the self-management process also includes helping patients become more knowledgeable and at the same time, assisting them in feeling empowered (ie, more confident in their skills to manage the illness) [24,37,38]. The participants in this study emphasized that obtaining knowledge and learning management skills through an eHealth intervention could be an accessible way of gaining knowledge, providing them with a constant option they could use at their own preferred time and pace, allowing for self-directed repeated exposure to the information and making the process familiar. Providing knowledge and support through the use of ICT could give patients the opportunity to reinforce knowledge and skills over time, possibly prolonging positive effects over time. In that sense, eHealth interventions could strengthen potentially fading effects of CBT on pain and function [9,39]. Existing research has supported this notion and found internet-based interventions to be a viable way for patients to obtain skills and knowledge, even after formal treatment is completed [12], showing long-term effects [16].

Providing information alone, however, is not sufficient if the goal is to provide pain management and behavioral change [24]. The participants in this study supported this notion, describing a need for a variety of content and functionality in order to become better at managing their own pain.

Self-Monitoring to Support Balance and Activity Pacing in Everyday Life

Balancing everyday life and pacing activities were some of the main topics emphasized. Although they were familiar with the concept of activity pacing, few participants actually practiced this activity. People with chronic pain sometimes avoid physical activity due to a fear of increased pain, but typically also push themselves too hard despite the experienced pain. Either approach will likely result in poorer overall functioning across time [7]. To enhance awareness of how activities and mood might be related to their pain, and support actual behavior change, the participants in this study suggested establishing daily registrations of variables such as sleep, mood, and activity. Keeping a medical diary for tracking one's own medical history and writing a daily pain journal were other suggestions. These suggestions are all related to awareness and potential for identifying patterns, helping patients become more aware of their own actions and behavior, which is in line with self-management and CBT goals as well as fostering of health behavior change. Electronic registrations and notes have also gained acceptance as methods for supporting patients in keeping a more reliable and up-to-date diary, instead of one based on recollection [40,41]. These findings are consistent with research identifying factors that determine the success and failure of eHealth interventions, indicating self-management and empowerment as the most important factors for successful outcomes [42]. Studies have also supported the notion of technology-based interventions leading to patient empowerment, as they can encourage patients to take more ownership over their personal health [43] and reduce patients' dependency on health care services [38].

Design Features to Support Motivation and Usage

Motivation to complete and continue the use of an eHealth intervention is crucial in order to obtain an effect. Participants in this study expressed interest in eHealth interventions for chronic pain, but despite such interest [19], previous studies reported high attrition and dropout rates for eHealth pain management interventions [44,45]. There is still a need for more information related to acceptance or intended use of eHealth interventions and to adherence [46]. In a study featuring an unguided eHealth pain management intervention, high acceptance did not result in high uptake or adherence [47]. Guided eHealth interventions appear to have less attrition and drop-out challenges [48]. Having a therapist following the treatment could be motivating, allowing for more personalized treatment and important feedback [48]. However, a recent review found only a small difference between self-guided and therapist-guided interventions in relation to drop-outs [15]. The participants in this study had different opinions on the value of or need for a therapist's support. Some of the participants expressed a wish for contact with health care providers throughout the intervention, particularly in case of arising questions. At the same time, several participants stated that they did not want health care professionals to follow their progression on tasks or exercises, as this could increase the pressure on them, enforcing their frequent and collective feeling of "never doing enough." All participants emphasized that they wanted the intervention to give them positive input and support, noting that they did not want an intervention that focused on their problems or reminded them of all the things they were not able to do. This supports the notion that close attention should be given to designing a positive user experience with persuasive and engaging features to trigger a positive effect and potentially promote adherence [49].

Examining potential facilitators and barriers for use of an eHealth tool in line with recommendations [42], we found that accessibility, usability, personalization, and reliability were factors that were emphasized by participants as important facilitators for use, supporting the existing literature [21,29]. Poor usability, including a flawed design and an overwhelming amount of information, was emphasized as key barriers for use. The participants emphasized the need for a tool they could use daily, independent of varying pain intensity, activity level, and concentration capacity. These results support previous findings where patients with chronic pain described high demands related to the design of Web-based pain management interventions, even higher than their physicians, indicating challenges for distracting content and texts that were too dense [50].

Strengths and Limitations

Several aspects of "trustworthiness" [51] were covered well in this study of qualitative research. The credibility was assured by presenting the steps in the analysis as thoroughly as possible as well as showing examples with quotes in the Results section. Transparency was assured by researcher triangulation and by presenting citations from the participants. Dependability was assured by describing the analytical process in detail in the study to make it possible for the reader to agree with and understand the logic of the findings.

This study had several limitations. First, only five spouses were included in this study, and the results might have been different if more spouses had participated. However, the patients were the main focus of the study. Spouses were included to gain additional perspectives on patients' needs, and adding their voice was seen as a study strength. Second, most of the patients participating were women, and the results may therefore not fully cover the male perspective of how eHealth interventions could support them. Third, most participants had lived with chronic pain for a long time, which may have influenced their description of everyday challenges and needs. Many had also participated in self-management courses, which meant that they already knew a lot about pain and pain management. As such, patients with a shorter history of chronic pain could have other needs and preferences not covered in this study.

Study participation may be due to higher motivation and engagement in self-management than the average patient with chronic pain. Most participants were interested in how they could learn to live and better cope with the pain, rather than how they could get rid of the pain, indicating acceptance and, perhaps, maturity in how this sample approached their pain. Finally, it is not unreasonable to assume that those willing to participate in the study were people with a special interest in technology, as half of the participants had already tried some forms of health-related apps/technology. The results might therefore not fully capture facilitators and barriers that apply to a less technology-experienced group. However, considering the pervasive use of technology in today's society, most patients with chronic pain likely have some form of ICT experience, and the clear requirement of accessibility and usability should be representative regardless of technology experience. Future studies should explore these topics further, examine how current eHealth interventions can adequately address known challenges with interventions (eg, adherence), and seek to incorporate existing and new findings into the design and development of new eHealth self-management interventions.

Conclusions

To our knowledge, this is the first study to explore the experiences of patients with chronic pain with regard to ICT, understand how an eHealth intervention can support the everyday needs and challenges of patients with chronic pain, and identify the possible facilitators and barriers for patients' use of an eHealth pain management intervention. The participants (ie, patients with chronic pain and their spouses) considered ICT an acceptable way of gathering self-management support, particularly emphasizing the need for more information and knowledge, finding a better balance in everyday life, and obtaining support for improved communication and social participation. The participants described an ideal eHealth intervention as one that could be used for self-management support and distraction from pain, at any time or in any situation, regardless of the varying pain intensity and concentration capacity. The results provide insight into the future potential of eHealth interventions aiming to support self-management for patients with chronic pain.

Acknowledgments

This study was supported by the Norwegian Research Council (Grant # 256574 to Principal Investigator LSN). The authors would like to thank the patients and spouses who participated in this study. The authors would also like to thank all collaborative partners, especially the partners at Vestre Viken Hospital trust and Drammen municipality for their help with participant recruitment, and occupational therapist Katherine Bevan, in particular, for her enthusiasm and support in the recruitment process and for making the interviews possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guides.

[PDF File (Adobe PDF File), 110KB - jmir_v21i4e13205_app1.pdf]

Multimedia Appendix 2

Additional quotes from participants (related to Tables 2 and 3).

[PDF File (Adobe PDF File), 110KB - jmir_v21i4e13205_app2.pdf]

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
2. Carnes D. Patterns of chronic pain in the population. International Journal of Osteopathic Medicine 2011 Sep 24;14(3):81-85. [doi: 10.1016/j.ijosm.2011.07.004]
3. Solberg Nes L, Roach AR, Segerstrom SC. Executive functions, self-regulation, and chronic pain: a review. Ann Behav Med 2009 Apr;37(2):173-183. [doi: 10.1007/s12160-009-9096-5] [Medline: 19357933]

4. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
5. Solberg Nes L, Carlson CR, Crofford LJ, de Leeuw R, Segerstrom SC. Self-regulatory deficits in fibromyalgia and temporomandibular disorders. Pain 2010 Oct;151(1):37-44. [doi: 10.1016/j.pain.2010.05.009] [Medline: 20561749]
6. Bergeson SC, Dean JD. A systems approach to patient-centered care. JAMA 2006 Dec 20;296(23):2848-2851. [doi: 10.1001/jama.296.23.2848] [Medline: 17179462]
7. Mann EG, Lefort S, Vandenkerkhof EG. Self-management interventions for chronic pain. Pain Manag 2013 May;3(3):211-222. [doi: 10.2217/pmt.13.9] [Medline: 24654764]
8. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999 Mar;80(1-2):1-13. [Medline: 10204712]
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. [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
10. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of Mobile Apps in Pain Management: Systematic Review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
11. Vugts MAP, Joosen MCW, van der Geer JE, Zedlitz AMEE, Vrijhoef HJM. The effectiveness of various computer-based interventions for patients with chronic pain or functional somatic syndromes: A systematic review and meta-analysis. PLoS One 2018;13(5):e0196467 [FREE Full text] [doi: 10.1371/journal.pone.0196467] [Medline: 29768436]
12. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. Internet Interv 2016 May;4:17-34 [FREE Full text] [doi: 10.1016/j.invent.2015.12.001] [Medline: 30135787]
13. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A Systematic Review of Technology-assisted Self-Management Interventions for Chronic Pain: Looking Across Treatment Modalities. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]
14. Wilson M, Roll JM, Corbett C, Barbosa-Leiker C. Empowering Patients with Persistent Pain Using an Internet-based Self-Management Program. Pain Manag Nurs 2015 Aug;16(4):503-514. [doi: 10.1016/j.pmn.2014.09.009] [Medline: 26088940]
15. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2018 Nov 01:1-19. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
16. Dear BF, Gandy M, Karin E, Fogliati R, Fogliati VJ, Staples LG, et al. The Pain Course: 12- and 24-Month Outcomes From a Randomized Controlled Trial of an Internet-Delivered Pain Management Program Provided With Different Levels of Clinician Support. J Pain 2018 Dec;19(12):1491-1503. [doi: 10.1016/j.jpain.2018.07.005] [Medline: 30099209]
17. European Comission. 2018. eHealth: Digital health and care URL: https://ec.europa.eu/health/ehealth/overview_en [accessed 2019-03-04] [WebCite Cache ID 76cQpp8lV]
18. Keogh E. Developments in the use of e-health for chronic pain management. Pain Manag 2013 Jan;3(1):27-33. [doi: 10.2217/pmt.12.70] [Medline: 24645929]
19. Schneider LH, Hadjistavropoulos HD. When in doubt, ask the audience: potential users' perceptions of Internet-delivered cognitive behavioural therapy for chronic pain. Pain Res Manag 2014;19(4):173-178 [FREE Full text] [Medline: 24911177]
20. McDermott MS, While AE. Maximizing the healthcare environment: a systematic review exploring the potential of computer technology to promote self-management of chronic illness in healthcare settings. Patient Educ Couns 2013 Jul;92(1):13-22. [doi: 10.1016/j.pec.2013.02.014] [Medline: 23566427]
21. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
22. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
23. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
24. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
25. Cranen K, Groothuis-Oudshoorn CG, Vollenbroek-Hutten MM, IJzerman MJ. Toward Patient-Centered Telerehabilitation Design: Understanding Chronic Pain Patients' Preferences for Web-Based Exercise Telerehabilitation Using a Discrete Choice Experiment. J Med Internet Res 2017 Dec 20;19(1):e26 [FREE Full text] [doi: 10.2196/jmir.5951] [Medline: 28108429]
26. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]

27. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
28. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
29. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an Internet-based chronic pain self-management program. Int J Med Inform 2017 Dec;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]
30. Brinkmann S, Kvale S. InterViews: learning the craft of qualitative research interviewing. Thousand Oaks, CA: Sage Publications; 2014.
31. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
32. Stickdorn M, Schneider J. This Is Service Design Thinking: Basics, Tools, Cases. Hoboken, NJ: Wiley; 2012.
33. Braun A, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis. International Journal of Qualitative Methods 2017 Oct 02;16(1):160940691773384-160940691773313. [doi: 10.1177/1609406917733847]
35. Macdonald GG, Koehn C, Attara G, Stordy A, Allerdings M, Leese J, et al. Patient Perspectives on the Challenges and Responsibilities of Living With Chronic Inflammatory Diseases: Qualitative Study. J Participat Med 2018 Nov 21;10(4):e10815. [doi: 10.2196/10815]
36. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
37. Creer TL, Holroyd KA. Self-management of chronic conditions: the legacy of Sir William Osler. Chronic Illn 2006 Mar;2(1):7-14. [doi: 10.1177/17423953060020010501] [Medline: 17175675]
38. Coster S, Norman I. Cochrane reviews of educational and self-management interventions to guide nursing practice: a review. Int J Nurs Stud 2009 Apr;46(4):508-528. [doi: 10.1016/j.ijnurstu.2008.09.009] [Medline: 19012889]
39. Henschke N, Ostelo RW, van Tulder MW, Vlaeyen JW, Morley S, Assendelft WJ, et al. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2010 Jul 07(7):CD002014. [doi: 10.1002/14651858.CD002014.pub3] [Medline: 20614428]
40. Dansie EJ, Turk DC. Assessment of patients with chronic pain. Br J Anaesth 2013 Jul;111(1):19-25 [FREE Full text] [doi: 10.1093/bja/aet124] [Medline: 23794641]
41. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 07;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
42. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
43. 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]
44. Macea DD, Gajos K, Daglia CYA, 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]
45. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
46. 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]
47. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925] [Medline: 30131313]
48. Andersson G. The Internet And Cbt: A Clinical Guide. Boca Raton, FL: CRC Press; 2014.
49. Ludden GD, van Rompay TJL, Kelders SM, van Gemert-Pijnen JEWC. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]

50. Gremeaux V, Coudeyre E, Givron P, Hérisson C, Pélissier J, Poiraudeau S, et al. Qualitative evaluation of the expectations of low back pain patients with regard to information gained through semi-directed navigation on the Internet. Ann Readapt Med Phys 2007 Jul;50(6):348-355. [doi: 10.1016/j.annrmp.2007.03.017] [Medline: 17513002]
51. Guba EG, Lincoln Y. Naturalistic Inquiry. Thousand Oaks, CA: SAGE Publications; 1985.

Abbreviations

CBT: cognitive behavioral therapy
eHealth: electronic health
ICT: information and communication technology

Edited by G Eysenbach; submitted 20.12.18; peer-reviewed by S Chokshi, S Donnelly, H Devan, J Moral-Munoz; comments to author 17.01.19; revised version received 20.02.19; accepted 25.02.19; published 01.04.19.

Please cite as:
Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, Haaland-Øverby M, Heldal K, Schreurs KMG, Waxenberg LB, Weiss KE, Morrison EJ, Solberg Nes L
Patients' Needs and Requirements for eHealth Pain Management Interventions: Qualitative Study
J Med Internet Res 2019;21(4):e13205
URL: https://www.jmir.org/2019/4/e13205/
doi:10.2196/13205
PMID:30877780

Original Paper

A Digital Game and School-Based Intervention for Students in Hong Kong: Quasi-Experimental Design

Angie KY Shum[1], MClinPsych; Eliza SY Lai[1], MSocSc; Wing Gi Leung[1], MSc; Mabel NS Cheng[1], MSocSc; Ho Kit Wong[1], BSc; Sam WK So[1], MPhil; Yik Wa Law[2], PhD; Paul SF Yip[1], PhD

[1]The Hong Kong Jockey Club Centre for Suicide Research and Prevention, The University of Hong Kong, Hong Kong, China (Hong Kong)
[2]Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Paul SF Yip, PhD
The Hong Kong Jockey Club Centre for Suicide Research and Prevention
The University of Hong Kong
2/F, The Hong Kong Jockey Club Building for Interdisciplinary Research
5 Sassoon Road, Pokfulam
Hong Kong,
China (Hong Kong)
Phone: 852 28315232
Fax: 852 25497161
Email: sfpyip@hku.hk

Abstract

Background: In Hong Kong, with an increasing number of children experiencing mental health issues, there is a need to not only develop innovative interventions but also develop comprehensive prevention interventions so as to reduce their anxiety symptoms and enhance their emotional management and interpersonal relationships.

Objective: The aim of this study was to determine the effectiveness of *The Adventures of DoReMiFa*, an integration model of the cognitive-behavioral approach and positive psychology by using digital game–based and school-based mental health enhancement intervention to magnify the social and emotional health and well-being of the school children in Hong Kong aged 9 to 11 years.

Methods: A quasi-experimental design method was used to evaluate this digital game and school-based intervention. *The Adventures of DoReMiFa* was piloted in 4 primary schools where students were allocated to either an intervention or a control group. The participants were assessed at pre- and postintervention with a 6-month follow-up measuring their mental health knowledge, levels of anxiety symptoms, positive and negative thinking, perspective-taking, and self-esteem.

Results: A total of 459 primary school students from 4 primary schools participated in the study. The response rate on the questionnaires answered on the Web was up to 85.1% (391/459). Compared with the control group, the intervention group was found to have significant association with improved mental health knowledge at the time immediately after the intervention (beta=.46; *P*=.01) and in the 6-month postintervention period (beta=.66; *P*<.001); for perspective-taking, the intervention group had exhibited a significant improvement 6 months after the completion of the universal program (beta=1.50; *P*=.03). The intervention, however, was found not to be effective in reducing the rates of anxiety symptoms and negative thinking among the participating students.

Conclusions: *The Adventures of DoReMiFa*, an integration of a digital game–based and school-based mental health enhancement intervention, was shown to be effective in elevating the knowledge of mental health and promoting perspective-taking in the primary school students of Hong Kong. Although there was insufficient evidence to support a reduction in symptoms of anxiety and negative automatic thoughts, the overall results were still encouraging in that a preventive effect was found, indicating that the program has the potential to enhance the mental well-being of schoolchildren. It also suggests that knowledge enhancement may not necessarily lead to behavior change, and more focused effort may be needed to achieve the translation. The implications and limitations of this study and suggestions for future research were also discussed.

(J Med Internet Res 2019;21(4):e12003) doi:10.2196/12003

KEYWORDS

digital game-based learning; school-based learning; mental health; schools; students; child welfare; health promotion; follow-up studies; internet access; public health; non-randomized controlled trials

Introduction

Childhood Mental Disorders

Mental health is defined as a state of well-being in which an individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community. [1]

The US Department of Health and Human Services further emphasizes that mental health in childhood is characterized by the achievement of developmentally appropriate milestones such as effective coping and healthy interpersonal skills [2]. Although there are various specific definitions of mental health in different cultures, together, the overall concepts incorporate the social, emotional, and behavioral well-being of children and adolescents. In younger children, common signs of distress may include crying and expressing worries and fear (emotional), isolating oneself (social) or aggressive behaviors, and withdrawing from peers and other pleasurable activities (behavioral) and the like. Although many children occasionally display these symptoms, it is the frequency and severity of those symptoms that are of concern.

Internalizing problems, specifically depression and anxiety disorders, are often characterized by covert symptomatology and are therefore difficult to recognize and verbalize [3]. Epidemiological studies have shown that the prevalence rate of specific internalizing difficulties in children and adolescents, that is, depression and anxiety disorders, has been alarmingly high in the past decades, and anxiety disorders have been shown to be the most common form of mental illnesses in the youngsters [4-6]. The prevalence rate of anxiety disorders in children varies from 10% to 23%, with the lifetime prevalence rate estimated to be about 29% [6-8]. The results from another national community survey further indicated that nearly 1 in 3 American youths aged 13 to 18 years has met the criterion of an anxiety disorder [9].

Prevalence of Childhood Anxiety in Hong Kong

In Hong Kong, the prevalence of internalizing difficulties among children and adolescents has drawn great attention to educators, clinicians, and researchers. In a prevalence study on childhood internalizing problems, Siu examined 1598 mothers of local primary school children and reported that the prevalence of internalizing difficulties among the young children was 11.4% [10]. However, the actual rate of prevalence may have been overlooked as many children may remain unidentified by their parents and teachers. In particular, owing to the strong stigmatization and discrimination in Chinese societies, despite the parents being aware of the illnesses in their children, they may tend to minimize the symptoms [11].

Expanding Mental Health Promotion in the School Setting

Although there are many effective treatments available for childhood anxiety disorders such as cognitive behavioral therapies, medication treatments, or a combination of the two [12-14], nonetheless, many children fail to receive timely and formal treatments. Furthermore, it has been shown that half of the mental disorders found in adults start at the age of 14 years, where anxiety symptoms among adults surface even earlier, with a median age of 11 years [6]. As a result, preventive intervention may seem to be an ideal method, focusing on young children before the onset of the symptoms and diagnosable disorders. Thus, school would be an ideal platform to promote the mental health of children. During the past decades, research found that schools have become the most common point of entry and provider of mental health promotion services. It is also shown to have various advantages, including reaching out to a larger number of students and reducing common obstacles to treatment access such as time and costs [15,16].

Over the past 10 years, the Hong Kong Jockey Club Centre for Suicide Research and Prevention (The Centre) has been implementing various mental health promotion programs for primary and secondary school students in Hong Kong. In 2006, based on the cognitive-behavioral theory, the first universal school-based prevention program, *Little Prince is Depressed*, was initially developed with the aim to reduce depressive symptoms and enhance protective factors of secondary school students in Hong Kong. The results indicated a significant improvement in help-seeking attitudes and self-esteem among the students who have completed the teacher-led mental health program [17]; however, there were also considerable challenges such as a small sample size, tight teaching schedule, and low teacher efficacy and so forth. To tackle the difficulties and limitations mentioned above, in 2010, an internet-based mental health program, *Professor Gooley and The Flame of Mind* [18], was further developed. It consisted of cognitive behavioral theory and positive psychology frameworks targeted to enhance the mental well-being of adolescents. Despite the innovative design and interesting storyline, the Web-based program also encountered problems such as high dropout and low completion rates.

Framework of This Study

Learning from past experiences, it is believed that the involvement of teachers in school programs can be decisive in motivating and engaging students to complete the programs. Furthermore, in the Web-based programs, the students are exposed to the contents of the program before each school-based lesson, which may aid to tackle the tight teaching schedule. This in turn provides more time for the teachers to guide interactive activities, for discussions and reflections during classes, and these are believed to be useful to consolidate the students' knowledge and the skills learned. With these benefits, this study aimed to investigate into the effectiveness of integrating both school-based and digital game–based frameworks. It is believed that all students can benefit from such a form of program delivery mode and skills building program. It is predicted that the intervention effect of this program may further be strengthened through the enhancement of the children's intrapersonal (such as self-awareness, self-esteem, and perspective-taking) and interpersonal (such as social and communication skills) functioning in schools. In addition to the elements of the cognitive-behavioral approach, this program also integrated key elements of the positive psychology perspectives to promote the mental well-being of students.

Positive psychology consists of valued subjective experiences such as well-being, contentment, hope, flow, and happiness. Its fundamental goal is to promote positive mental health in the community [19,20]. In recent years, positive education programs have become increasingly popular in schools, with an ultimate goal to enhance the well-being of students and parents as well as teachers. Various studies have been conducted to demonstrate the potential benefits of the positive education intervention programs in cultivating positive emotions, behaviors, and attitudes among the schoolchildren [21,22]. Given the significant impacts of positive education among children, key elements of the positive education model, that is, positive emotions, empathy, and gratitude, were integrated into this program. It is believed that enhanced perspective-taking skills and the exercise of gratitude could yield significant improvement in the children's well-being. In addition, although there is no formal curriculum for mental health education in Hong Kong, some of its concepts may be brought up during homeroom or moral and civic education lessons. Nevertheless, there is still a lack of evidence-based and systematic mental health education programs for young children in the schools in Hong Kong.

Furthermore, the use of the internet has already penetrated through people's daily lives. Many studies have shown that the internet has been increasingly utilized by health care providers to provide the latest information and interactions with their patients and the general public, in particular, among the young generation [23-26]. In addition, in view of the high-demanding curriculum and competitive teaching hours at school, a transition to a more digital learning environment seems to be necessary for the students in Hong Kong. As such, to attract the students' interest in mental health and promote their learning in the relevant skills, a digital game–based approach was designed and integrated into the entire program. It was written into an adventurous story, and the purpose of each digital game–based lesson was used to prepare the students for the content of the school-based lesson.

Aims of This Study

In Hong Kong, the majority of studies to date have focused on treatment interventions. This study however sought to extend the research to explore into the prevention of childhood anxiety disorders and the enhancement of the students' mental well-being through an innovative online and offline approach. The objective was to examine the effectiveness of implementing a school-based and digital game–based intervention program with a combination of the cognitive-behavioral and positive psychology model. Through this, it aimed to reduce childhood anxiety and improve the mental well-being of students. This study has 4 specific objectives: (1) the primary objective was to compare the self-reported anxieties in an intervention condition and a monitoring condition at different time points. It was hypothesized that compared with the children in the monitoring group, the intervention group would have greater reduction in self-reported anxiety after completing the program; (2) the second objective was to examine the effectiveness of the interventions on the children's mental health knowledge. It was hypothesized that the children in the intervention condition would evidence a greater enhancement in the relevant knowledge; (3) the third objective was to examine the children's mental well-being. The intervention group was expected to have more positive thinking, enhanced self-esteem, and empathy skills at the postintervention and the 6-month follow-up time intervals; and (4) the final objective was to evaluate the effectiveness of digital game–based learning in equipping students with the relevant knowledge and skills. It was hypothesized that compared with the students who completed less on the Web-based learning, students who had a high completion rate (at least completed 6 modules) on the Web-based learning would result in lower rates of self-reported anxiety symptoms and better mental well-being.

Methods

Recruitment

Intervention Effect

As randomization was not preferred by the participating schools, this study is a quasi-experimental design. On the basis of their needs, suitability, and capability, the school teachers would prefer to assign all students of the same grade (ie, either Primary 4 or Primary 5) to either an intervention or a control condition. All participating schools had both intervention and control groups (ie, students of 1 grade was the intervention group and another grade in the same school was the control group). Before the baseline assessment, informed consent was obtained from all participants and their parents. The students were asked to fill out questionnaires at the preintervention stage (T0), postintervention stage, that is, 2 weeks after completion of the program (T1), and at the 6-month follow-up (T2).

Digital Game–Based Learning Effect

On the basis of the same group of students recruited above, further analyses were conducted to assess the efficacy of the digital game–based program. On the basis of their completion rate on the digital game–based learning, the students were divided into 2 groups. Those students who have reached the completion rate of 50% or above were in the high completion group, whereas those below 50% were in the low completion group.

Procedure

In total, 4 primary schools completed the program in the period of 2014 to 2015. At the enrollment stage, an information seminar on the program was organized in June 2014. After the seminar, a few schools showed their interest to join the program. Invitation letters including information sheets, the program outline, and enrollment forms were then sent to the interested schools and all primary schools in Hong Kong. An initial consent was obtained from the principal of each school to invite the students, their parents, and teachers to participate in this study. On the basis of the needs, suitability, and capability of the students, the school teachers would assign the students of 1 grade as an intervention group and another grade as a control group. All the parents of the students were sent an information sheet describing the program, and an informed consent form was to be completed and returned by the parents. The consent rate was 42.50% (459/1080).

The Adventures of DoReMiFa was a digital game–based and school-based mental health enhancement program for Primary 4 and 5 students aged 8 to 12 years. The content was designed and developed by multidisciplinary professionals including clinical psychologists, social workers, counselors, and educators and adopted the cognitive-behavioral therapeutic approach and positive psychology as theoretical frameworks. The program consisted of 8 modules including (1) emotional competence; (2) cognitive model; (3) ABC theory; (4) problem-solving skills; (5) social skills; (6) communication skills; (7) empathy; and (8) gratitude. The program was a combination of digital game–based and school-based teaching. Altogether, the 8 modules were transformed into 11 digital game–based lessons and 8 classroom teachings. This project was supported by the Quality Education Fund of the Education Bureau of Hong Kong.

To prepare for the design of the digital game–based lessons, a few focus groups were conducted with the students, parents, and teachers separately to collect their views on mental health, the program design, and digital game learning. The digital game–based lessons were written into an adventurous story that combines the elements of a storyline, dialogues, problem-solving, challenges, mini-games, teamwork, and so on. The story was about 4 monsters, Do, Re, Mi, and Fa, who came from another planet to look for a book hidden in a primary school. The hidden book would help to improve the mental health of their fellow citizens. They would encounter 2 students, Lily and Max, after they landed on Earth to help them look for the book and solve different challenges and tasks. Each digital game–based lesson lasted about 20 min and was used to prepare the students for the content of the upcoming classroom lesson. Details of the program can be found at the official website of *The Adventures of DoReMiFa* [27]. Each student would be given a unique login name and password by their teacher during the school-based introduction lesson.

After each digital game–based lesson, a classroom lesson would follow to facilitate and consolidate the students' learning. The classroom lessons were led by graduates or students of a master's degree in Counseling or Counseling and Clinical Psychology or were qualified teachers who had received an 8-hour preservice training. All instructors would teach the same class throughout the 8 school-based lessons, and the class teachers were required to stay in the classroom to observe the teaching and facilitate the activities and discussions. The structure of each lesson (approximately 25 to 60 min/lesson) involved different interactive activities such as role-play and card games. Textbox 1 presents the corresponding digital game–based and school-based lessons in sequential order.

Textbox 1. Distribution of the digital game–based lessons and school-based teachings.

School-based lessons:

- Lesson 1: Introduction of the program
- Lesson 2: Emotional competence
- Lesson 3: Cognitive model and ABC theory
- Lesson 4: Problem-solving skills
- Lesson 5: Social and communication skills
- Lesson 6: Empathy
- Lesson 7: Gratitude
- Lesson 8: Review

Digital game–based lessons:

- Lesson 1: Emotional competence
- Lesson 2-3: Cognitive model and ABC theory
- Lesson 4-5: Problem-solving skills
- Lesson 6-8: Social and communication skills
- Lesson 9: Empathy
- Lesson 10: Gratitude
- Lesson 11: Review

In this study, the students were assessed at 3 time intervals—before the first digital game–based lesson (T0: preintervention), 2 weeks after completion of the program (T1: postintervention), and 6 months after completion of the program (T2: follow-up). The questionnaire was self-administered and consisted of 5 parts to measure 5 outcomes of the study, including (1) anxiety; (2) mental health knowledge; (3) positive and negative thinking; (4) perspective-taking; and (5) self-esteem. All participants in the intervention group and control conditions were invited to complete the same set of questionnaires online through the official website of *The Adventures of DoReMiFa* in a similar time frame. Informed consent was sent to all students and their parents before the study and all students could opt out from the study at any time. Background information of the students such as class numbers and dates of birth were used as identifiers to match with the postintervention and follow-up tests.

Measurements

The Screen for Child Anxiety–Related Emotional Disorders

The Screen for Child Anxiety–Related Emotional Disorders (SCARED) is designed to evaluate anxiety disorder symptoms of children and adolescents aged 8 to 18 years [28]. It consists of 41 items measuring 5 anxiety disorders. In this study, 9 items were used to measure the children's Generalized Anxiety Disorder on a 3-point Likert scale (from 0=Not True or Hardly Ever True to 2=Very True or Often True). The optimal cutoff for the Generalized Anxiety Disorder subscale was 8. The whole SCARED scale had a good internal consistency (Cronbach alpha=.95) as well as its subscales (Cronbach alpha=.72 to .88) and was a suitable and useful instrument to screen for anxiety disorders in children and adolescents in a community or school setting [29-31]. A Chinese version of the scale was used in this study.

Mental Health Knowledge Checklist

An 11-item knowledge checklist was designed by the research team to examine the mental health knowledge of the students before and after the program. Each item on the checklist was generated from the knowledge and skills taught in the program, and all items were *True* or *False* questions. Students would score 1 point for a correct answer and 0 for an incorrect answer. A Chinese version of the scale was used in this study.

Children's Automatic Thoughts Scale-Negative or Positive

The original Children's Automatic Thought Scale-Negative/Positive (CATS-N/P) consists of 4 subscales, each has 10 items measuring negative self-statements across both the internalizing and externalizing problems of children and adolescents aged 8 to 17 years [32]. In this study, only 1 subscale, that is, personal failure, was adopted to assess the children's negative thoughts about themselves. In addition to the personal failure subscale, another 10 positive items were added to facilitate the calculation of the state-of-mind ratios [33]. Then, corresponding to the frequency of their experience over the past week, the students were asked to rate each statement on a 5-point Likert scale from *Not at all* (0) to *All the time* (4). Higher scores reflect a greater number of negative or positive automatic thoughts. As indicated by Cronbach alpha coefficient of .79 to .95, CATS-N/P has good internal consistency and test-retest reliability [32,34]. A Chinese version of the scale was used in this study.

Interpersonal Reactivity Index

The original Interpersonal Reactivity Index (IRI) is a 28-item self-report measure of the students' perspective-taking behaviors [33]. In this study, a Chinese version of the IRI (C-IRI) was adopted [35]. The items were grouped under a 3-factor model—personal distress, fantasy scale, and empathy. The empathy subscale (11 items) mostly consists of items taken from the perspective-taking and empathic concern subscales of the original IRI. In this study, the perspective-taking items (6 items) from the empathy subscale in the C-IRI were used. Furthermore, the students were asked to rate each statement on a 5-point Likert scale from 0 (Does not describe me well) to 4 (Describes me very well). Higher scores show a higher level of perspective-taking (empathy). The C-IRI was found to have acceptable psychometric properties in the Chinese context with good internal consistency (Cronbach alpha of .65 to .70) and a 2-week test-retest reliability of .68 to .83 [35].

Rosenberg Self-Esteem Scale

The Rosenberg Self-Esteem Scale (RSES) consists of 10 items measuring the students' self-esteem [36]. All items were answered using a 4-point Likert scale format ranging from *Strongly disagree* (1) to *Strongly agree* (4). After reversing the scores of 5 negative items, higher scores indicate a higher level of self-esteem. This scale has good reliability and validity. Fischer and Corcoran [37] showed that the 2-week test-retest reliability was high (rs>.80) and the scale correlated significantly with other measures of self-esteem and depression and anxiety in predicted directions. The Chinese version of RSES was validated by Yeung [38] and used in this study.

Data Analyses

1. Intervention Effect

Multilevel modeling was used to analyze the data. Similar analytical approach was employed in an evaluation study on a school-based prevention program of depression [39]. A 3-level regression model was first employed to test if significant group differences (between the intervention group and the control group) existed during the pretest assessment on each outcome measure including knowledge, anxiety, negative and positive automatic thoughts, perspective-taking, and self-esteem.

To test if intervention effects existed between the 2 groups, a series of 4 level models were individually employed on each outcome. Level 1 accounted for the changes within the students on the outcome measure, whereas levels 2, 3, and 4 accounted for changes between students, between classes, and between schools, respectively. Independent variables included gender, age, time variable (T0 as the pretest [ie, baseline], T1 as the posttest, and T2 as the follow-up test), the group variable (intervention and control), and their interaction term (ie, time×group). The interaction term *T1* × group examined the differences across 2 groups immediately after completion of the program. A statistically significant coefficient indicated the changes of scores varied across groups, suggesting an existence of interaction effects. Similarly, the interaction term *T2* × group examined the differences across groups during the follow-up assessment and was employed to test whether there was an interaction effect after 6 months of completion of the program.

If significant interaction effects were found in both *T1* × and *T2* × groups, this would indicate significant persistence of intervention effects, suggesting effects were found immediately and have persisted for 6 months after completion of the program. If, however, significant interaction effects were only found in the *T1* × group but not in the *T2* × group, this would indicate significant immediate intervention effects but diminish afterward. On the contrary, if significant interaction effects were only found in the *T2* × group but not in the *T1* × group, this would indicate no intervention effect right after completion of the program but a significant 6-month delayed intervention effect.

2. Digital Game–Based Learning Effect

Further analyses on the intervention group were conducted to determine the effects of Web-based programs in terms of the number of Web-based modules completed at the end of the implementation of the program. There were 11 modules in total. On the basis of the number of completed modules, the students from the intervention group were further categorized into 2 groups (below and above the completion rate of 50%, that is, whether completed at least 6 modules). Multilevel analyses, as described, were then used to determine the completion levels of the Web-based modules in relation to the improvement in outcome measures within the intervention group at different assessment intervals. A significant coefficient of the interaction term time×group would suggest that if sufficient modules were completed by the students, they would have experienced relatively greater improvement in a certain outcome measure.

All analyses in 1. and 2. above were performed by PROC MIXED in the statistical program SAS (Statistical Analysis System) version 9.3 for Windows by SAS institute [40], and all statistical significances were determined at a 5% level.

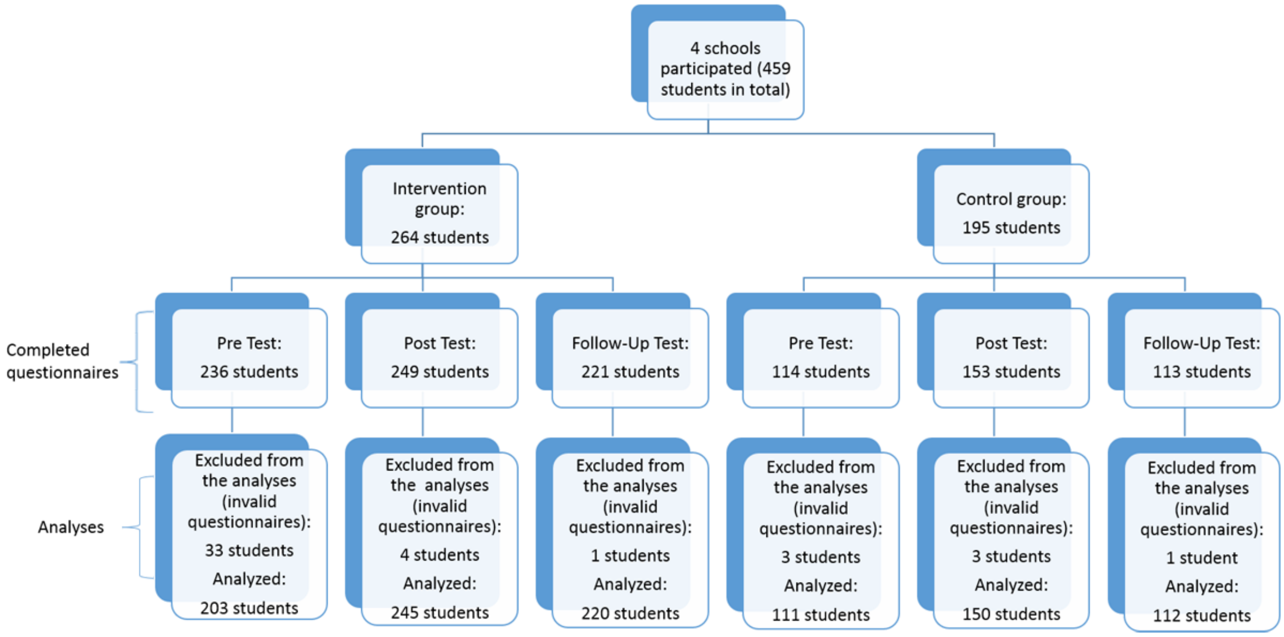
Results

Intervention Effect

Demographic Information of the Participants

The participants were 459 children aged 8 to 12 years from 4 primary schools in Hong Kong. It included 3 coeducational schools and 1 girls' school. Overall, 264 and 195 students were in the intervention group and control group, respectively; 232 were Primary 4 students (50.6%; 232/459) and 227 (49.4%; 227/459) were Primary 5 students. The mean age of the students was 9.53 years with an SD of 0.717 in the intervention group and 9.48 years in the control group (SD 0.64). Entries by students who showed the same response for all items across 3 measures or above and those who only completed 1 out of the 3 time points were excluded from the analyses. In the end, 314, 395, and 332 students were put into the analyses in T0, T1, and T2, respectively. Please refer to Figure 1 for the flow of data exclusion.

Figure 1. Completed questionnaires analyses. Participants who only completed one assessment point and questionnaires with sloppy answers were excluded from the analyses. As some students did not initially complete the pretest questionnaires, more participants completed the posttest questionnaires than the pretest questionnaires in the analyses but managed to complete both post and follow-up tests.



Differences at the Pretest Assessment

At the pretest assessment, no significant differences were found between the 2 groups across knowledge, anxiety, negative automatic thoughts, perspective-taking, and self-esteem. Positive automatic thoughts, however, were found to have a significant difference during baseline assessment across the 2 groups.

Multimedia Appendix 1 summarizes the scores of all outcomes measured during pretest, posttest, and follow-up assessments between the intervention and control groups.

Results of the estimated interaction effects from multilevel modeling are shown in Multimedia Appendix 2. Significant results were found in the participants' mental health knowledge and perspective-taking (empathy). The intervention effects were strong in mental health knowledge. A significant intervention effect was found immediately after completion of the program. The students showed significant enhanced mental health knowledge in the intervention group (beta=.46; *P*=.01). Follow-up assessments showed a very prominent persistent effect of intervention where further improvement on mental health knowledge was found after 6 months of completion of the program (beta=.66; *P*<.001). As for perspective-taking, no significant intervention effect was found immediately after completion of the program; however, a significant delayed intervention effect was found at T2, 6 months after completion of the program (beta=1.50; *P*=.03). No significant intervention effects were found in the outcome measures of anxiety, positive and negative automatic thoughts, and self-esteem among the participants immediately and 6 months after completion of the program. Although there was a slight increase of the mean score in the outcome measure of anxiety level at T2 of the intervention group, all scores were within the normal range. Although no intervention effects were found, on the whole, positive improvements and trends were shown by the means at each time point of negative and positive automatic thoughts and self-esteem.

Digital Game–Based Learning Effect

Demographic Information of the Participants

A total of 264 students from the intervention group were divided into high and low completion groups. Of the 11 modules, 182 (68.9%; 182/264) students reached the completion rate of 50% or above (high completion group), whereas 82 (31.1%; 82/264) completed 50% or below (low completion group). Overall, 139 (76.4%; 139/182) females and 43 (23.6%; 43/182) males were in the high completion group, whereas 43 (52.4%; 43/82) females and 39 (47.6%; 39/82) males were in the low completion group.

No significant differences were found on each outcome measure between the students of the high and the low completion of modules at the pretest assessment (the third far right column), which suggest that students' participation in the Web-based modules was not likely to be associated with their pretest performances.

Multilevel modeling was then applied to identify the existence of any significant differences in the improvement on each outcome measure between students with higher and lower completion of Web-based modules since pretest, at posttest, and follow-up assessments. Results of the estimated effects are given in Multimedia Appendix 2 (the far right 2 columns). When compared with the lower completion group (ie, completion rate below 50%), among the higher completion group (ie, completion rate above 50%), and immediately after completion of the program, significant improvements were found on the students' mental health knowledge (beta=.51; *P*=.04) and positive automatic thoughts (beta=3.32; *P*=.007). However, no significant differences in improvement between the higher and lower completion groups were found on the 6 outcome measures 6 months after completion of the program.

Discussion

Principal Findings

This study showed that the combination of school-based and digital game–based program, *The Adventures of DoReMiFa*, was effective in promoting the students' mental health knowledge, positive automatic thoughts, and perspective-taking in Primary 4 and 5 students in Hong Kong. Specifically, when compared with the control group, the results showed an immediate and sustained enhancement on the students' mental health knowledge after implementation of the program. Also, the program was likely to act as a buffer against deterioration of the students' empathetic skills (perspective-taking) 6 months after completion of the program, indicating a delayed intervention effect on this protective factor. Additional exploration was also conducted to investigate into the impact of the high versus low completion rate of the digital game–based intervention program among students who were intervened. When compared with those in the low completion rate group, it was found that students in the high completion rate group showed better results in mental health knowledge and enhanced positive automatic thoughts immediately after implementation of the program; however, a sustained effect was not found.

The study also provided sufficient evidence on the effectiveness of utilizing both digital game–based and school-based methods in delivering mental health enhancement programs in Hong Kong. It was shown to be particularly effective in the mental health knowledge of the intervened students even 6 months after implementation of the program, suggesting a prominent continued effect of enhancement in mental health knowledge among the students. This further indicated that the students had advanced their understanding of various concepts and the skills taught in the program, including emotional competence, problem-solving, communication skills, empathy, cognitive model, ABC theory, and so on. Furthermore, the results of this study also found a delayed intervention effect on perspective-taking, a measure on empathy. The effect, however, was only found to be prominent 6 months after implementation of the program but not immediately. Besides, it was surprising that no instant or sustained effects were found in empathetic skills; however, when compared with the control group, the students in the control group showed a significant deterioration of perspective-taking in 6 months. This phenomenon may imply that this program could act as a buffer toward the reduction of empathy among the students in the intervention group, that is, students in the intervention group maintained their level of empathy throughout the program. Another explanation is that the taught skill, empathy, may require longer duration to be nurtured among the students. In other words, it is possible that not only the students might need extra time to understand the concept but also time to transfer it into practice, which then in turn may enhance their thorough understanding on empathy. No significant effects, negative automatic thoughts, and self-esteem, however, were found in the measures of anxiety and well-being. Possible explanations could be the students in this study were not randomly assigned into intervention and control groups, and also, the duration of the intervention period was short. Therefore, future studies could look into the optimal intervention duration that best benefits the students in Hong Kong and make further improvement on the research method for this type of study to be more effective.

Moreover, the project team intended to investigate into the changes among the students on each individual outcome, and although not many tests were conducted, significant results were found. Multiple comparisons were further conducted to investigate into the prominence of the proposed outcome measures. In both the program intervention and digital game–based learning effect, among the known significant effects of mental health knowledge, perspective-taking, and positive automatic thoughts, after the program implementation, a relatively strong sustained intervention effect was observed on the mental health knowledge of the Primary 4 and 5 students. In addition, a strong significance on positive automatic thoughts was also observed in the students who were in the high completion rate group of the digital game–based program. Thus, this further shows the robustness of the effectiveness of such a digital game–based and school-based mental health enhancement program.

Although numerous studies have shown that universal school-based programs are effective in yielding positive outcomes to reduce anxiety-related symptoms in children and adolescences [41,42], this study failed to show a significant reduction in anxiety among the intervened students. One explanation is that the program was entirely led by mental health professionals rather than by school teachers. Previous research studies have shown positive impacts on the program effectiveness in reducing anxiety symptoms when mental health programs were led by teachers rather than by mental health professionals. Recently, a local study conducted in The Centre by Lai et al [43] also found that teacher-led mental health prevention programs yielded better outcomes in reducing anxiety symptoms in students than professional-led groups. It was thus suggested that, in the long-run, classroom teachers will achieve better sustainability in school-based mental health prevention programs as they have sufficient amount of time with the students and already have pre-existing rapport built throughout the school year [44]. Hence, in view of these research studies, with the aim to deliver the best quality program to promote well-being and reduce anxiety and depressive symptoms, future studies should utilize classroom teachers as school-based mental health prevention program instructors. Another possible explanation could be due to this program being a universal program where the mean scores of the students' anxiety levels were within the normal range before the implementation program; therefore, it is of no surprise that there were no significant reductions of this outcome measure after the program implementation. Similar results were also found in a previous study suggesting that the students who had more depressive symptoms before the program implementation showed better improvement in cognitive restructuring skills and support-seeking behaviors after the program [17]. Thus, future studies may investigate into the intervention effect between the students who score higher in the anxiety level before the program and in the control study. All the above limitations mentioned indicate that perhaps the implementation of a mental health program solely in 1 school year may not be effective in improving the students' mental health if the school environment remains unchanged. With the new era of the whole school approach and implementation of mental health programs in the school curriculum [45], it is of utmost importance to promote the whole school approach on the mental well-being among schools. Hence, it is worthwhile investigating into the effectiveness of delivering a high-quality mental health program in the school curriculum. Through this investigation and implementation, the research team anticipates yielding positive and promising significant results in improving the mental health of the students in Hong Kong.

Moreover, this program exhibits positive impacts and effectiveness in that it has provided more evidence on the effectiveness of the universal mental health enhancement program in the school setting. In addition, moving onto a preventive mental health model, many overseas researchers have been investigating into the effectiveness of increasing the social, emotional, and psychological skills in the students via mental health enhancement programs in school. A great example conducted by Seligman et al [46] has brought tremendous impacts on the students by using positive psychology in the school setting. This study further showed this positive phenomenon through the application of positive psychology as well as cognitive behavioral theory in schoolchildren. All these studies successfully showed promising results in supporting and nurturing the mental health of students in a holistic approach and under the classroom setting [46-48]. In addition, this study further suggested that schools can be a great platform for students to learn about mental health–related knowledge. Furthermore, primary school students encounter many ups and downs within the school setting, suggesting that school-based programs on mental health promotion designed according to the school context may be beneficial to the students, as this would assist them in learning and applying the taught techniques and theories to their daily situations.

To the best of the authors' knowledge, thus far, no study has investigated into the effectiveness of utilizing the mode of delivery in conjunction with school-based and digital game–based interventions to implement mental health school programs in Hong Kong. Recent evidence showed promising results on the efficacy of computer- and internet-based cognitive behavioral treatments among the youth with anxiety and depression, indicating that online-based mode of delivery could be an alternative to traditional face-to-face treatments [13]. Moreover, previous research on school-based or digital game–based programs has also demonstrated their own plausible proficiency and effectiveness. The former showed decreased levels of depression, anxiety and stress, and enhanced mental health knowledge and positive attitudes toward mental illnesses among the adolescents, whereas the latter showed good multimedia capabilities, far-reaching abilities, and timeless and effortless accessibilities [39,49]. Nonetheless, a major limitation was found on the attrition of users while implementing the digital game–based program alone. A possible explanation is that the program was a self-motivated learning program where students were asked to complete without the involvement of teachers. It was reported that an overall 57% attrition rate was found in the systematic review and meta-analysis of the digital game–based program [50], indicating that dropouts are common during the implementation of the internet-based program and thus suggesting that it may not fully excel when implemented on its own. However, the attrition rate can be reduced when the programs were supported by therapists or with additional administrative support, demonstrating that human support will enhance the effectiveness of such programs. To address the engagement rate of the digital game–based programs, this study has incorporated mental health professionals as the human support of the program. The results revealed that the intervened students in the high completion group showed significantly higher scores in mental health knowledge and positive automatic thoughts than the students in the low completion group. This implied that a positive impact was found where a higher engagement rate of the digital game–based program was linked to better outcomes of the students' knowledge on mental health and the thinking processes. Moreover, with the fast-growing pace of the internet in the past decades, delivering a mental health program as a digital game–based intervention may in fact be a cost-effective intervention in promoting mental health. This is so as this program can be delivered on the medium of the internet with no additional costs despite the increased number of beneficiaries. Furthermore, the digital game platform can be sustained and maintained on the internet server with no extra costs, meaning that the Web-based program can be used continuously after the program implementation. Also, as stated by the World Health Organization [51], depression and anxiety have cost a significant economic impact on the global economy, where it is believed that good utilization of the internet intervention may perform part of the preventive work on anxiety and depression, in turn minimizing the economic cost due to mental illnesses and unemployment. Nevertheless, e-engagement work is still challenging, requiring much effort and research into understanding how it can most benefit the students. Future studies should therefore focus on investigating into what additional components should be added into the program to enhance efficacy, adherence, and engagement of such modes of delivery.

Limitations

Few limitations of this study were identified for future research and direction. First, no significant improvements were found in the outcome measures of anxiety, negative automatic thoughts, and self-esteem. The possible reasons could be the program was not led by teachers, had limited and inconsistent research time for questionnaire completion, and also, the teaching time was short, thus limiting the students' learning skills and consolidation. Furthermore, it can be observed that solely implementing a mental health program may not be effective in promoting mental health and reducing the symptoms of anxiety in the students. Second, despite further analyses of the low and high digital game–based completion levels being compared between the 2 groups, the conducted statistical analyses may not specifically reveal the effectiveness of the digital game–based element of the program. Possible explanation of the results could be due to the integration model of the school- and digital game–based programs. Third, utilizing both digital game–based and school-based modes of delivery indeed showed a better engagement rate of the online platform rather than solely implementing the digital game–based program alone; however, over a quarter of the students showed low engagement in the digital game–based program. This also indicated that there was a lack of investigation into the students' acceptability and confidence toward the presented mode of program delivery. Fourth, this study is a quasi-experimental design where a limitation would be the lack of randomization between the control and intervention groups. Finally, this study only gathered 1 posttest immediately after the program implementation, where the long-term and sustainability effects of such a program were not captured.

Future Direction

Future research can aim to conduct a longitudinal study to best capture the effects of school-based and digital game–based mental health programs across a longer period of time, with consistent and longer class periods across various schools. In addition, further investigation will be needed to look into the optimal intervention duration that will best benefit the students such as the number of school-based and digital game–based lessons. Also, with the aim to enhance program quality and provide continuous support to the students, future programs should consider training classroom teachers to lead the program instead of employing mental health professionals. Furthermore, it has been suggested to give the students sufficient amount of time to complete the questionnaires to best capture any changes in the outcome. In response to the above constraints from the schools, future studies should focus on implementing mental health programs in the school curriculum to allow higher quality and more dosage of mental health programs in the schools in Hong Kong. To assess the effectiveness of digital game–based intervention, future studies should also include an intervention group with students who solely engage in the Web-based component of the program. With regard to enhancing the engagement of the digital game–based intervention among the students, the research team will further look into additional components that would reduce the attrition rate of students so as to increase the efficacy, adherence, and engagement of such a mode of program delivery. Moreover, adopting a similar approach conducted by Poder et al [52], a thorough investigation will be conducted to access the students' acceptability and confidence level toward the presented mode of program delivery. In relation to the randomization between the control and intervention groups, future research in this area may incorporate a propensity score matching method into the studies. As already mentioned, the long-term effect and sustainability of the effectiveness of the program will be examined in future studies where more than 1 posttest will be conducted to examine the persistence of knowledge and skills gained.

Conclusions

In this study, *The Adventures of DoReMiFa*, an innovative and integrated model of a school-based and digital game–based mental health promotion program, used as a universal preventive strategy, was shown to be effective in the enhancement of the students' mental health knowledge and the maintaining of the standard of perspective-taking. This further demonstrated that early intervention and preventive programs for mental health promotion in schools were an effective way in educating the students on the importance of good mental well-being. This study also reinforces that such programs and practices should disseminate across various schools in Hong Kong so as to promote the well-being of the students and administer adaptive skills in their daily encounters. Nonetheless, to achieve a behavior change, in addition to knowledge gain, more effort may be needed. Furthermore, at times, other complimentary efforts are needed to make some standalone programs more effective.

Acknowledgments

This study was funded by the Quality Education Fund (Project No. 2012/0048), Education Bureau of the Government of the Hong Kong Special Administrative Region from November 2013-July 2015.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Pretest, posttest, and follow-up mean scores (SDs) by group, and tests of the intervention effects from multilevel modeling. Remarks: 1 - There was an inconsistent number of students (N) across each outcome measure. Students who completed 2 time points or more on each outcome measure were put into the analyses. 2 - According to Figure 1, a total 314 students' information was in T0, which is inconsistent with the number of students shown in this table. On top of the 314 students who were put into the analyses, the remaining samples were those who did not complete T0, but completed T1 and T2.

[PDF File (Adobe PDF File), 46KB - jmir_v21i4e12003_app1.pdf]

Multimedia Appendix 2

Pretest, posttest, and follow-up mean scores (SDs) by the number of completed Web-based modules among the students in the intervention group, and tests of the intervention effects from multilevel modeling. Remarks: 1 - There was an inconsistent number of students (N) across each outcome measures. Students who completed 2 time points or more on each outcome measure were put into the analyses. 2 - According to Figure 1, a total of 203 students' information was in T0, which is inconsistent with the number of students shown in this table. On top of the 203 students who were put into the analyses, the remaining samples were those who did not complete T0, but completed T1 and T2.

[PDF File (Adobe PDF File), 47KB - jmir_v21i4e12003_app2.pdf]

References

1. World Health Organization. Mental health: strengthening our response URL: http://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2018-11-08] [WebCite Cache ID 73lTB7c04]
2. Perou R, Bitsko RH, Blumberg SJ, Pastor P, Ghandour RM, Gfroerer JC, Centers for Disease Control and Prevention (CDC). Mental health surveillance among children--United States, 2005-2011. MMWR Suppl 2013 May 17;62(2):1-35. [Medline: 23677130]
3. Zahn-Waxler C, Klimes-Dougan B, Slattery MJ. Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. Dev Psychopathol 2000;12(3):443-466. [doi: 10.1017/S0954579400003102] [Medline: 11014747]
4. Donovan CL, Spence SH. Prevention of childhood anxiety disorders. Clin Psychol Rev 2000 Jun;20(4):509-531. [doi: 10.1016/S0272-7358(99)00040-9] [Medline: 10832552]
5. Cartwright-Hatton S, McNicol K, Doubleday E. Anxiety in a neglected population: prevalence of anxiety disorders in pre-adolescent children. Clin Psychol Rev 2006 Nov;26(7):817-833. [doi: 10.1016/j.cpr.2005.12.002] [Medline: 16517038]
6. 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]
7. Dadds MR, Spence SH, Holland DE, Barrett PM, Laurens KR. Prevention and early intervention for anxiety disorders: a controlled trial. J Consult Clin Psychol 1997 Aug;65(4):627-635. [Medline: 9256564]
8. Muris P, Merckelbach H, Mayer B, Prins E. How serious are common childhood fears? Behav Res Ther 2000 Mar;38(3):217-228. [Medline: 10665156]
9. Merikangas KR, He J, Burstein M, Swendsen J, Avenevoli S, Case B, et al. Service utilization for lifetime mental disorders in US adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2011 Jan;50(1):32-45 [FREE Full text] [doi: 10.1016/j.jaac.2010.10.006] [Medline: 21156268]

10. Siu AF. A prevalence study on internalizing problems among primary school children in Hong Kong. J Child Fam Stud 2008 Jan 23;17(6):779-790. [doi: 10.1007/s10826-008-9189-y]
11. Tsang HW, Tam PK, Chan F, Cheung W. Stigmatizing attitudes towards individuals with mental illness in Hong Kong: Implications for their recovery. J Community Psychol 2003 Jun 05;31(4):383-396. [doi: 10.1002/jcop.10055]
12. Compton SN, Peris TS, Almirall D, Birmaher B, Sherrill J, Kendall PC, et al. Predictors and moderators of treatment response in childhood anxiety disorders: results from the CAMS trial. J Consult Clin Psychol 2014 Apr;82(2):212-224 [FREE Full text] [doi: 10.1037/a0035458] [Medline: 24417601]
13. 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;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
14. James A, Soler A, Weatherall R. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev 2005 Oct 19(4):CD004690. [doi: 10.1002/14651858.CD004690.pub2] [Medline: 16235374]
15. Farmer EM, Burns BJ, Phillips SD, Angold A, Costello EJ. Pathways into and through mental health services for children and adolescents. Psychiatr Serv 2003 Jan;54(1):60-66. [doi: 10.1176/appi.ps.54.1.60] [Medline: 12509668]
16. Ruini C, Belaise C, Brombin C, Caffo E, Fava GA. Well-being therapy in school settings: a pilot study. Psychother Psychosom 2006;75(6):331-336. [doi: 10.1159/000095438] [Medline: 17053333]
17. Wong PW, Fu K, Chan KY, Chan WS, Liu PM, Law Y, et al. Effectiveness of a universal school-based programme for preventing depression in Chinese adolescents: a quasi-experimental pilot study. J Affect Disord 2012 Dec 15;142(1-3):106-114. [doi: 10.1016/j.jad.2012.03.050] [Medline: 22840612]
18. The HKJC Centre for Suicide Research and Prevention. Professor Gooley and the Flame of Mind URL: http://www.gooley.edu.hk/login.php [accessed 2018-11-08] [WebCite Cache ID 73lUSAHbn]
19. Norrish JM, Williams P, O'Connor M, Robinson J. An applied framework for positive education. Int J Wellbeing 2013;3(2):147-161. [doi: 10.5502/ijw.v3i2.2]
20. Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000 Jan;55(1):5-14. [Medline: 11392865]
21. Shoshani A, Steinmetz S, Kanat-Maymon Y. Effects of the Maytiv positive psychology school program on early adolescents' well-being, engagement, and achievement. J Sch Psychol 2016 Dec;57:73-92. [doi: 10.1016/j.jsp.2016.05.003] [Medline: 27425567]
22. Froh JJ, Sefick WJ, Emmons RA. Counting blessings in early adolescents: an experimental study of gratitude and subjective well-being. J Sch Psychol 2008 Apr;46(2):213-233. [doi: 10.1016/j.jsp.2007.03.005] [Medline: 19083358]
23. Beck F, Richard JB, 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;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
24. Beckjord EB, Finney RL, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007 Jul;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
25. Culjak G, Kowalenko N, Spranca M, Tennant C. Internet Health Promotion and Behavior Change Theory: an Integrated Model Applied to Internet Interventions to Enhance Prevention and Early Detection. In: 2015 48th Hawaii International Conference on System Sciences. 2015 Presented at: 48th Hawaii International Conference on System Sciences; 5-8 Jan, 2015; Kauai, HI, USA p. - URL: https://ieeexplore.ieee.org/document/7070174 [doi: 10.1109/HICSS.2015.359]
26. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
27. Quality Education Fund. The Adventures of DoReMiFa URL: https://www.doremifa.edu.hk/ [accessed 2018-11-08] [WebCite Cache ID 73lUhF1kT]
28. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry 1997 Apr;36(4):545-553. [doi: 10.1097/00004583-199704000-00018] [Medline: 9100430]
29. Chan SM, Leung CH. Factor Structure of the Screen for Child Anxiety-Related Emotional Disorders (SCARED) in a Community Sample of Hong Kong Chinese Adolescents. Child Psychiatry Hum Dev 2015 Oct;46(5):671-682. [doi: 10.1007/s10578-014-0509-8] [Medline: 25288523]
30. Hale WW, Crocetti E, Raaijmakers QA, Meeus WH. A meta-analysis of the cross-cultural psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED). J Child Psychol Psychiatry 2011 Jan;52(1):80-90. [doi: 10.1111/j.1469-7610.2010.02285.x] [Medline: 20662993]
31. Su L, Wang K, Fan F, Su Y, Gao X. Reliability and validity of the screen for child anxiety related emotional disorders (SCARED) in Chinese children. J Anxiety Disord 2008 May;22(4):612-621. [doi: 10.1016/j.janxdis.2007.05.011] [Medline: 17628391]
32. Schniering CA, Rapee RM. Development and validation of a measure of children's automatic thoughts: the children's automatic thoughts scale. Behav Res Ther 2002 Sep;40(9):1091-1109. [Medline: 12296494]

33. Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol 1983;44(1):113-126. [doi: 10.1037/0022-3514.44.1.113]
34. Hogendoorn SM, Wolters LH, Vervoort L, Prins PJ, Boer F, Kooij E, et al. Measuring negative and positive thoughts in children: an adaptation of the Children's Automatic Thoughts Scale (CATS). Cognit Ther Res 2010 Oct;34(5):467-478 [FREE Full text] [doi: 10.1007/s10608-010-9306-2] [Medline: 20930924]
35. Siu AM, Shek DT. Validation of the interpersonal reactivity index in a Chinese context. Res Soc Work Pract 2016 Aug 18;15(2):118-126. [doi: 10.1177/1049731504270384]
36. Rosenberg M. Conceiving the Self. USA: Krieger Pub Co; 1986.
37. Fischer J, Corcoran K. Measures for Clinical Practice and Research: A Sourcebook Volume 2: Adults. New York: Oxford University Press; 2007.
38. Yeung K. University of Hong Kong. 1998. The dynamics of interparental conflict and adolescent's behavior problems URL: https://kaching.socialwork.hku.hk/Share/Scales/Self-Esteem%20Result.pdf [accessed 2019-03-01] [WebCite Cache ID 76YEs00Yc]
39. Sawyer MG, Pfeiffer S, Spence SH, Bond L, Graetz B, Kay D, et al. School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative. J Child Psychol Psychiatry 2010 Feb;51(2):199-209. [doi: 10.1111/j.1469-7610.2009.02136.x] [Medline: 19702662]
40. SAS Institute. 2011. SAS/IML® 9.3 User’s Guide URL: http://support.sas.com/documentation/cdl/en/imlug/64248/PDF/default/imlug.pdf [accessed 2019-03-01] [WebCite Cache ID 76YF3rEcu]
41. Corrieri S, Heider D, Conrad I, Blume A, König H, Riedel-Heller SG. School-based prevention programs for depression and anxiety in adolescence: a systematic review. Health Promot Int 2014 Sep;29(3):427-441. [doi: 10.1093/heapro/dat001] [Medline: 23376883]
42. Neil AL, Christensen H. Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. Clin Psychol Rev 2009 Apr;29(3):208-215. [doi: 10.1016/j.cpr.2009.01.002] [Medline: 19232805]
43. Lai ES, Kwok C, Wong PW, Fu K, Law Y, Yip PS. The effectiveness and sustainability of a universal school-based programme for preventing depression in Chinese adolescents: a follow-up study using quasi-experimental design. PLoS One 2016;11(2):e0149854 [FREE Full text] [doi: 10.1371/journal.pone.0149854] [Medline: 26921275]
44. Chodkiewicz AR, Boyle C. Positive psychology school-based interventions: a reflection on current success and future directions. Rev Educ 2016 Aug 30;5(1):60-86 [FREE Full text] [doi: 10.1002/rev3.3080]
45. Lewallen TC, Hunt H, Potts-Datema W, Zaza S, Giles W. The whole school, whole community, whole child model: a new approach for improving educational attainment and healthy development for students. J Sch Health 2015 Nov;85(11):729-739 [FREE Full text] [doi: 10.1111/josh.12310] [Medline: 26440815]
46. Seligman ME, Ernst RM, Gillham J, Reivich K, Linkins M. Positive education: positive psychology and classroom interventions. Oxford Rev Edu 2009 Jun;35(3):293-311 [FREE Full text] [doi: 10.1080/03054980902934563]
47. Waters L. A review of school-based positive psychology interventions. The Australian Educational and Developmental Psychologist 2012 Mar 27;28(02):75-90. [doi: 10.1375/aedp.28.2.75]
48. Schonert-Reichl KA, Oberle E, Lawlor MS, Abbott D, Thomson K, Oberlander TF, et al. Enhancing cognitive and social-emotional development through a simple-to-administer mindfulness-based school program for elementary school children: a randomized controlled trial. Dev Psychol 2015 Jan;51(1):52-66 [FREE Full text] [doi: 10.1037/a0038454] [Medline: 25546595]
49. Huen JM, Lai ES, Shum AK, So SW, Chan MK, Wong PW, et al. Evaluation of a digital game-based learning program for enhancing youth mental health: a structural equation modeling of the program effectiveness. JMIR Ment Health 2016 Oct 07;3(4):e46 [FREE Full text] [doi: 10.2196/mental.5656] [Medline: 27717921]
50. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
51. World Health Organization. 2017. Mental health in workplace URL: https://www.who.int/mental_health/in_the_workplace/en/ [WebCite Cache ID 74EgOZX6G]
52. Poder TG, Bellemare CA, Bédard SK, Lemieux R. Social acceptance and population confidence in telehealth in Quebec. BMC Health Serv Res 2015 Feb 21;15:72 [FREE Full text] [doi: 10.1186/s12913-015-0727-1] [Medline: 25889230]

Abbreviations

CATS-N/P: Children’s Automatic Thoughts Scale-Negative or Positive
C-IRI: Chinese-Interpersonal Reactivity Index
IRI: Interpersonal Reactivity Index
RSES: Rosenberg Self-Esteem Scale
SCARED: Screen for Child Anxiety–Related Emotional Disorders

Edited by G Eysenbach; submitted 23.08.18; peer-reviewed by A Amresh, D Katz, T Poder; comments to author 04.10.18; revised version received 28.11.18; accepted 03.02.19; published 05.04.19.

Please cite as:
Shum AKY, Lai ESY, Leung WG, Cheng MNS, Wong HK, So SWK, Law YW, Yip PSF
A Digital Game and School-Based Intervention for Students in Hong Kong: Quasi-Experimental Design
J Med Internet Res 2019;21(4):e12003
URL: https://www.jmir.org/2019/4/e12003/
doi:10.2196/12003
PMID:30950795

Review

Technology-Based Alcohol Interventions in Primary Care: Systematic Review

Alex T Ramsey[1], PhD; Jason M Satterfield[2], PhD; Donald R Gerke[3], PhD; Enola K Proctor[4], PhD

[1]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, United States
[2]Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[3]Graduate School of Social Work, University of Denver, Denver, CO, United States
[4]Brown School of Social Work, Washington University in St Louis, St Louis, MO, United States

Corresponding Author:
Alex T Ramsey, PhD
Department of Psychiatry
Washington University School of Medicine
660 South Euclid, Campus Box 8134
St Louis, MO, 63110
United States
Phone: 1 3143625370
Email: aramsey@wustl.edu

Abstract

Background: Primary care settings are uniquely positioned to reach individuals at risk of alcohol use disorder through technology-delivered behavioral health interventions. Despite emerging effectiveness data, few efforts have been made to summarize the collective findings from these delivery approaches.

Objective: The aim of this study was to review recent literature on the use of technology to deliver, enhance, or support the implementation of alcohol-related interventions in primary care. We focused on addressing questions related to (1) categorization or target of the intervention, (2) descriptive characteristics and context of delivery, (3) reported efficacy, and (4) factors influencing efficacy.

Methods: We conducted a comprehensive search and systematic review of completed studies at the intersection of primary care, technology, and alcohol-related problems published from January 2000 to December 2018 within EBSCO databases, ProQuest Dissertations, and Cochrane Reviews. Of 2307 initial records, 42 were included and coded independently by 2 investigators.

Results: Compared with the years of 2000 to 2009, published studies on technology-based alcohol interventions in primary care nearly tripled during the years of 2010 to 2018. Of the 42 included studies, 28 (64%) were randomized controlled trials. Furthermore, studies were rated on risk of bias and found to be predominantly low risk (n=18), followed by moderate risk (n=16), and high risk (n=8). Of the 24 studies with primary or secondary efficacy outcomes related to drinking and drinking-related harms, 17 (71%) reported reduced drinking or harm in all primary and secondary efficacy outcomes. Furthermore, of the 31 studies with direct comparisons with treatment as usual (TAU), 13 (42%) reported that at least half of the primary and secondary efficacy outcomes of the technology-based interventions were superior to TAU. High efficacy was associated with provider involvement and the reported use of an implementation strategy to deliver the technology-based intervention.

Conclusions: Our systematic review has highlighted a pattern of growth in the number of studies evaluating technology-based alcohol interventions in primary care. Although these interventions appear to be largely beneficial in primary care, outcomes may be enhanced by provider involvement and implementation strategy use. This review enables better understanding of the typologies and efficacy of these interventions and informs recommendations for those developing and implementing technology-based alcohol interventions in primary care settings.

(J Med Internet Res 2019;21(4):e10859) doi:10.2196/10859

KEYWORDS

alcohol drinking; risky health behavior; alcohol-related disorders; internet; computers; mobile health; primary health care; implementation science; review

Introduction

Background

Alcohol use is a leading risk factor for global disease burden, and recent findings indicate that even light-to-moderate drinking is detrimental to all-cause mortality [1,2]. Additionally, approximately 14% of US adults annually, and nearly 30% for lifetime, engage in harmful drinking consistent with alcohol use disorder [3]. This disorder contributes to over 200 diseases and health problems, including cirrhosis, cancers, fetal alcohol syndrome, assaults, and crash-related fatalities, with costs totaling US $249 billion annually in the United States alone [4-6]. Primary care settings that integrate physical and behavioral health care are uniquely situated to reach this at-risk population through delivery of evidence-based interventions (EBIs) to reduce harmful alcohol use [7,8].

To enhance the capacity for delivering behavioral health services, new approaches—such as those provided by digital technologies—can assess and intervene to reduce alcohol use (and associated harm) and facilitate referrals to specialty treatment [9]. In fact, health centers are increasingly leveraging technology to reduce medical staff burden, facilitate electronic health record (EHR) integration, improve standardization and fidelity, and enhance service efficiency [10,11]. Additionally, recent randomized controlled trials have begun to highlight the promise of using technology-supported platforms—including computers, kiosks, or tablets—to deliver efficacious alcohol-related interventions in primary care [7,11,12]. Relatedly, telephone-delivered interventions, although less novel, remain popular approaches in primary care, despite a lack of synthesized research on effectiveness and optimal implementation.

Systematic reviews exist for generalized alcohol interventions in primary care [13,14] and technology-based behavioral health interventions in other settings [15-17]. However, despite emerging effectiveness data in both adult and adolescent populations [17], few efforts have been made to summarize the collective findings of technology-based alcohol interventions in primary care settings. In total, 2 excellent Cochrane reviews on the effectiveness of brief alcohol interventions were recently published; one was based in primary care but did not focus on technology-based interventions [18] and the other focused on technology-based interventions but the focus was largely outside of primary care settings [19]. Another systematic review examined digital and computer-based alcohol interventions in primary care [20]; however, this review includes a broader set of technology-based interventions, addresses important effect modifiers, and provides a substantial update, adding 4 years (2015 to 2018) beyond the previous review. Use of technology in primary care is rapidly evolving, and systematically updating the collective knowledge gained from recent efforts in this area is needed to inform future delivery of technology-based alcohol interventions in these settings.

Furthermore, although technology-based behavioral health interventions are likely to be diverse in nature, there is currently a lack of conceptual clarity and no system for categorizing these interventions (eg, patient-facing, provider-facilitated, or a combination of both), making it difficult to directly compare interventions with similar purposes and approaches. The conceptualization and categorization provided here is useful for intervention developers and researchers in this area of investigation. The overarching purpose of this study was to review the most recent literature on the use of technology to deliver, enhance, or support the implementation of alcohol-related interventions in primary care.

Research Questions

In this review, we focused on addressing the following key questions regarding technology-based alcohol interventions in primary care:

1. What proportion of technology-based alcohol interventions is delivered (1) directly to the patient, (2) by the provider via technology-based medium, (3) some combination of both, or (4) directly to the provider to improve care?
2. What factors supported the use of technology-based alcohol interventions (technological platforms, delivery contexts, implementation strategies, and EHR integration)?
3. What proportion of studies reported that the technology-based intervention (1) reduced drinking or alcohol-related harms and (2) demonstrated superiority to treatment as usual (TAU)?
4. Did efficacy differ by the (1) context in which the intervention was delivered, (2) type of technology, (3) categorization or target of the technology-based intervention, and (4) use of an implementation strategy?

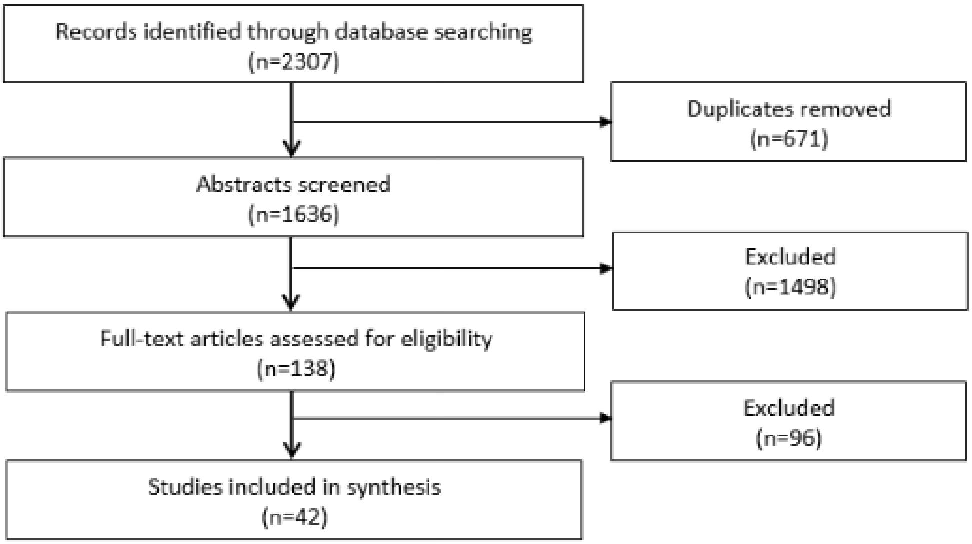
Answers to these questions would provide researchers with a better system of categorizing a diverse set of technology-based interventions and an understanding of the efficacy and effect modifiers of these interventions. These important contributions would inform recommendations for those developing and implementing technology-based alcohol interventions in primary care settings.

Methods

Search Strategy

In line with the Preferred Reporting Items for Systematic reviews and Meta-Analyses Statement and supporting publications to enhance the rigor of systematic reviews [21,22], we conducted a systematic literature review of English-language publications on completed research studies from January 2000 to December 2018. In total, 4 EBSCO databases (CINAHL Plus, Global Health, MEDLINE, and PsycInfo), ProQuest Dissertations, and Cochrane Reviews were searched. Although a comprehensive search of gray literature was not feasible, we did include dissertations and search trial registries (eg, PROSPERO). A total of 77 Boolean search terms were used to identify articles at the intersection of primary care, technology, and alcohol-related problems (see Multimedia Appendix 1). In conjunction with 2 university research librarians, extensive testing was conducted to limit the number of articles outside the inclusion criteria and yet ensure that the search strategy yielded comprehensive results. The systematic review protocol was published and is accessible on PROSPERO, an international prospective register of systematic reviews.

Figure 1. Flow diagram of search and screening results.



Study Selection

As illustrated in Figure 1, the initial database search yielded 2307 records, which reduced to 1636 records after removing duplicates. One of the 2 study investigators trained in conducting systematic reviews (AR or DG) reviewed the title and abstract of each record to assess study eligibility, and full text was obtained when appropriate. Studies were excluded if the article title or abstract did not specify that the study included a technology-based intervention that focused on alcohol use and was delivered in a primary care setting. A total of 138 studies were identified for full-text review and final selection for data abstraction.

Both study investigators (AR and DG) then independently reviewed each full text article for final inclusion, achieving an interrater reliability of .84. All initial discrepancies were resolved mutually through discussion between the same 2 investigators. Studies were excluded during this step if they were a self-defined pilot study or only reported on feasibility outcomes, alcohol use outcomes were not reported separately from other health outcomes, the so-called intervention only constituted screening for alcohol misuse, or the study results had previously been reported elsewhere (eg, main study outcome article). On the basis of this full-text review, 42 studies met our final inclusion criteria for coding [12,23-63].

Data Abstraction and Analysis

All data from included articles were recorded using a standardized data abstraction form, which both study investigators (AR and DG) completed independently, again mutually resolving all initial discrepancies through discussion. To address our key questions, articles were coded on a range of topics, including study design and sample size, type of technology, category or target of intervention, location of delivery, implementation strategies used, primary and secondary outcomes, intervention efficacy results, and a risk of bias score to inform study quality (see Multimedia Appendix 2). The wide heterogeneity of primary outcomes and assessment tools precluded the ability to conduct a meta-analysis; instead, data from the articles were primarily summarized descriptively. However, we conducted one-way analyses of variance (ANOVAs) with an efficacy score treated as the dependent variable and several of the factors listed above—location of delivery (ie, delivery context), type of technological platform used, category of intervention, and presence of an implementation strategy—treated as the independent variables.

Results

Descriptive Analyses

Of the 42 included studies, 5 (12%) studies were published from 2000 to 2004, 7 (17%) were published from 2005 to 2009, 21 were (50%) published from 2010 to 2014, and 9 (21%) were published from 2015 to 2018. The included studies were conducted in 8 different countries, with 33 of 42 studies (79%) based in the United States. Of the 42 included studies, 28 (64%) featured randomized designs, 10 (24%) were quasi-experimental, and 4 (10%) were observational studies. Multimedia Appendix 2 presents further information on designs for each study.

Type of Technology

Technology-based interventions were also categorized by the technological platform used to deliver the intervention. Of the 47 technological platforms identified, 13 (28%) were telephone or telehealth, 13 (28%) were stand-alone computer or software, 10 (21%) were Web-based, 5 (11%) were mobile (eg, tablet and smartphone), 3 (6%) were interactive voice response, 2 (4%) were kiosk, and 1 (2%) was video.

Category or Target of Intervention

Using a typology informed by research on behavior change at multiple levels [64], technology-based interventions were categorized into 4 main types: (1) *Patient-facing*; (2)

Provider-facilitated; (3) *Patient-facing plus provider-facilitated*; and (4) *Provider-directed*. Table 1 summarizes these types of technology-based interventions, including the conceptual definitions and representative examples of each type. The coding team determined that, of the 42 included studies, 14 (33%) interventions were patient-facing, 11 (26%) were provider-facilitated, 11 (26%) were patient-facing plus provider-facilitated, and 6 (14%) were provider-directed.

Delivery Context

Articles were coded into 3 broad delivery contexts: in clinic (eg, waiting room and exam room), out of clinic (eg, home and work), or both in and out of clinic (eg, part in waiting room and part at home). Of the 42 included articles, 18 (43%) reported interventions delivered in clinic, 17 (40%) reported interventions delivered at home (or otherwise outside of the clinic setting), and 7 (17%) reported interventions delivered both in clinic and at home.

Implementation Strategies

Implementation strategies, conceptualized as “methods or techniques used to enhance the adoption, implementation, and sustainability of a clinical program or practice” [65], constitute an important component to delivery of evidence-based practices [66,67], including those facilitated by the use of technology [68,69]. Previous research has categorized implementation strategies into the following categories: planning, educating, financing, restructuring, managing quality, and attending to the policy context [70]. Although underreported and underspecified in clinical research [65], implementation strategies are necessary to maximize the translation of research-based interventions into practice settings [71,72]. For this review, we only considered approaches to be implementation strategies when health care professionals (rather than research staff) were involved in efforts to improve uptake or delivery of the intervention. Of the 42 studies included, 15 (36%) specified the use of an implementation strategy to support the delivery of technology-based alcohol interventions (see Multimedia Appendix 2).

We documented 17 total and 11 unique implementation strategies using the Expert Recommendations for Implementing Change compilation [71]. The reported implementation strategies included the following: *Conduct ongoing training* (n=3), *Make training dynamic* (n=3), *Provide ongoing consultation* (n=2), *Remind clinicians* (n=2), *Relay clinical data to providers* (n=2), *Facilitation* (n=2), *Develop educational materials* (n=1), *Organize clinician implementation team meetings* (n=1), *Preparepatients/consumers to be active participants* (n=1), *Develop a formal implementation blueprint* (n=1), *Assess for readiness* (n=1), and *Conduct cyclical small tests of change* (n=1). These strategies varied widely in type and intensity, ranging from in-depth training on use of the technology-based intervention to the use of reminder cards and posters for providers. Implementation strategies also included providing frequent supervision, comparison between health care staff versus self (patient)-referral to the technology-based intervention and linking clinical management to the technology-based platform. Of note, only 6 of 42 articles (14%) mentioned any type of integration between the technology-based intervention and existing EHR systems.

Risk of Bias

We used a common classification scheme [73] to rate risk of bias pertaining to selection (eg, allocation concealment), performance (eg, blinding), detection (eg, validity of outcome assessment), attrition (eg, withdrawal rates), and reporting (eg, selective outcome reporting). We judged each of the 5 types of bias to be high (2), unclear (1), or low (0), and then calculated a sum risk of bias score on a scale of 0 to 10. Studies were then determined to have low (0 to 1), moderate (2 to 4), or high (5 to 10) overall risk of bias. Of the 42 studies, 18 (43%) were rated to be low risk of bias, 16 (38%) moderate risk of bias, and 8 (19%) high risk of bias.

Primary and Secondary Outcomes

We classified primary and secondary efficacy outcomes into 7 categories: quantity of alcohol use (eg, number of drinks per drinking day), frequency of alcohol use (eg, total number of drinking days), severity of alcohol use or risk scores (eg, Alcohol Use Disorders Identification Test or Alcohol, Smoking and Substance Involvement Screening Test score), binge or heavy episodic drinking (eg, number of binge episodes in the past week), status of at-risk alcohol use (eg, proportion of individuals with categorically defined at-risk drinking), any use (proportion of individuals with any past 90-day alcohol use), and drinking consequences (academic or legal problems related to drinking). Multimedia Appendix 3 organizes results across these categories of outcomes for each of the 4 technology types.

Table 1. Types of technology-based alcohol interventions.

Type	Conceptualization	Examples
Patient-facing	Intervention is delivered directly to the patient via technology with very limited or no provider involvement	Stand-alone touchscreen kiosk-based brief intervention
Provider-facilitated	Intervention is delivered by provider to patient via technology-based medium	Telephone-based brief intervention
Patient-facing plus provider-facilitated	Packaged intervention that has at least one patient-facing component and at least one provider-facilitated component	Tablet-based screening and brief intervention plus telephone-based counseling
Provider-directed	Intervention is delivered to provider to improve or support patient care delivery	Web-based training and clinical management dashboard

Efficacy of Intervention

Multimedia Appendix 3 summarizes study results according to statistical significance on 2 key outcomes—whether the intervention reduced drinking or drinking-related harms and whether the benefits were superior in comparison with TAU—for each primary and secondary outcome of each study. Of note, several studies reported multiple primary and secondary outcomes. At the study level, we assessed the proportion of primary and secondary efficacy outcomes that indicated reduced harm and that were determined to be superior to TAU. Similar methods of operationalizing and summarizing intervention efficacy have been used in previous systematic reviews in lieu of meta-analytic procedures [74].

Of the 24 studies with primary or secondary efficacy outcomes related to drinking and drinking-related harms, all 24 (100%) indicated reduced drinking or harm in at least half of the primary and secondary outcomes and 17 (71%) indicated reduced drinking or harm in all of the primary and secondary outcomes. Of the 31 studies with direct comparisons with TAU, 16 (52%) indicated that none of the primary and secondary outcomes were superior to TAU. However, of these 31 studies, 13 (42%) indicated that at least half of the primary and secondary outcomes were superior to TAU, and 8 (26%) indicated that all the primary and secondary outcomes were superior to TAU.

In examining predictors of intervention efficacy, we used the outcome of whether at least half of the primary and secondary efficacy outcomes were determined to be superior to TAU to maximize variability in the outcome, use a sufficiently rigorous cutoff, and focus on studies with direct comparisons with TAU. Of the 31 studies comparing intervention to TAU, 6 of 16 (38%) low risk of bias studies, 5 of 13 (38%) moderate risk of bias studies, and 2 of 2 high risk of bias studies reported at least half of their primary and secondary efficacy outcomes to be superior to TAU. This outcome was uncorrelated with the risk of bias (*P*=.812) variable.

Predictors of Efficacy

Descriptive analyses and one-way ANOVAs were used to examine intervention efficacy based on (1) whether the intervention was delivered in the clinic, at home, or both in the clinic and at home, (2) the type of technology used (computer or Web, telephone or video, and mobile), (3) the category or target of intervention (eg, patient-facing), and (4) the specification of an implementation strategy.

1. *Delivery context.* Of the 31 studies comparing intervention to TAU, 3 of 14 (21%) studies in the clinic, 5 of 10 (50%) studies at home, and 5 of 7 (71%) studies both in the clinic and at home showed at least half of the outcomes superior to TAU. The level of intervention efficacy did not differ significantly based on whether the intervention was conducted in the clinic (mean 0.21 (SD 0.43)), at home (mean 0.50 (SD 0.53)), or both in the clinic and at home (mean 0.71 (SD 0.49); $F_{2,28}$=2.81; *P*=.077). However, it should be noted here that there were only 7 cases in the category of both in the clinic and at home.
2. *Type of technology*. Of the 34 technological platforms within the 31 studies comparing intervention to TAU, 6 of 17 (35%) studies of computer or Web-based interventions, 7 of 13 (54%) telephone or video-based interventions, and 2 of 4 (50%) mobile-based interventions showed at least half of the outcomes superior to TAU. The level of intervention efficacy did not differ significantly based on whether the technology-based alcohol intervention was delivered via computer or Web (mean 0.35 (SD 0.49)), telephone or video (mean 0.54 (SD 0.52)), or mobile (mean 0.50 (SD 0.58); $F_{2,31}$=0.51; *P*=.603).
3. *Category of intervention.* The first analysis examined the level of intervention efficacy for each separate category of intervention; the second analysis examined the level of intervention efficacy between interventions that were only patient-facing versus those that had a provider-based component. Of the 31 studies comparing intervention with TAU, 2 of 12 (17%) patient-facing, 6 of 9 (67%) provider-facilitated, 5 of 9 (56%) patient-facing plus provider-facilitated, and 0 of 1 provider-directed studies showed at least half of their intervention outcomes superior to TAU. There were no significant differences in intervention efficacy between the individual categories of patient-facing (mean 0.17 (SD 0.39)), provider-facilitated (mean 0.67 (SD 0.50)), patient-facing plus provider-facilitated (mean 0.56 (SD 0.53)), and provider-directed, yet the results trended toward significance such that the patient-facing interventions showed lower efficacy than the other groups; ($F_{3,27}$= 2.54; *P*=.078). Indeed, when collapsing the groups that included a provider component and comparing them with patient-only interventions, we found that intervention efficacy was significantly higher for interventions that had a provider-based component (mean 0.58 (SD 0.51)) than for those that were patient-facing only (mean 0.17 (SD 0.39); $F_{1,29}$=5.76; *P*=.023, adjusted R^2=0.14 [medium effect size]).
4. *Specification of implementation strategy*. As implementation strategies were conceptualized as methods to improve the delivery of a clinical intervention, the provider-directed technologies were considered synonymous and indistinguishable from implementation strategies. Implementation strategies were analyzed as effect modifiers for the other types of interventions (eg, patient-facing); therefore, it was determined inappropriate to include the provider-directed technologies in this analysis as the intervention would have been the same as the effect modifier. Therefore, of the 31 studies comparing intervention with TAU, the provider-directed study (n=1) was removed for this particular analysis. In the remaining 30 studies, 6 of 6 (100%) studies reporting use of an implementation strategy showed at least half of the outcomes superior to TAU versus 7 of 24 (29%) studies not reporting use of an implementation strategy. Intervention efficacy was significantly higher when an implementation strategy was employed to facilitate delivery of the intervention (mean 1.00 (SD 0.00)) than for those with no specified implementation strategy (mean 0.29 (SD 0.46); $F_{1,28}$=13.60; *P*=.001; adjusted R^2=0.30 [large effect size]).

Discussion

Principal Findings

The use of technology-facilitated interventions in primary care settings is a burgeoning issue in behavioral health; however, research-based guidance is needed to inform development and implementation to ensure that these tools enhance, rather than impede, the efficiency and effectiveness of alcohol interventions in this setting. Our review attends to a number of key factors that may influence effectiveness of these interventions. Specifically, our review suggests a benefit to involving a provider in the delivery process, as compared with technology-based alcohol interventions that only engage the patient. This aligns with much research on technology-based interventions for use of tobacco and other substances [75,76]. Extensive research has shown that the specification and use of implementation strategies improve outcomes such as the adoption, reach, and sustainability of interventions [77-80]. However, to the extent that studies reported the use of implementation strategies when they were used, our review is among the first to find that employing an implementation strategy may actually enhance the effectiveness of a behavioral health intervention.

Our systematic review also highlights a pattern of growth in the number of studies evaluating technology-based alcohol interventions in primary care, with nearly 3 times as many studies on technology-based alcohol interventions in primary care published during the years of 2010 to 2018, as compared with the years of 2000 to 2009. The increasing number of studies on these interventions reflects an important angle of the changing health care landscape. As innovative technology-based approaches to delivering alcohol interventions continue to rapidly develop, it is necessary to take stock of the existing efforts and identify areas for further growth and improvement. Results indicated robust potential of technology-based interventions to support alcohol-related behavior change, with the majority indicating reduced drinking or harm in all of the reported primary and secondary outcomes. Similarly, when compared directly with TAU, there appeared to be strong efficacy for technology-based alcohol interventions over and above nontechnology-based alcohol intervention, with 42% (13/31) of studies with direct comparisons with TAU reporting that at least half of the primary and secondary efficacy outcomes of the technology-based interventions were superior to TAU.

Our review also highlights factors that appear to influence intervention efficacy. For instance, we identified 4 broad categories of technology-based alcohol interventions, and results indicated that interventions with a provider-based component (particularly provider-facilitated and patient-facing plus provider-facilitated interventions) were more efficacious than those that were patient-facing only. These findings suggest that provider involvement in the delivery of technology-based alcohol interventions may boost efficacy; however, further research is needed in this area.

Another key finding was that studies describing use of an implementation strategy reported more effective technology-based alcohol interventions. Although this effect has been documented in limited previous research [81], this remains a relatively novel finding that contributes to the accumulating evidence of the value-added benefit of employing implementation strategies to facilitate EBIs. Nevertheless, this finding is consistent with theoretical advances in the implementation science field that reject the assumption of *voltage drop* as an intervention moves from efficacy trials to real-world implementation studies [82]. Instead, it is reasonable to expect that with active strategies to adapt and tailor interventions to contexts and patients, the *voltage* of a technology-based intervention may even be enhanced. We encourage further empirical study of this potential effect in future research.

Yet, the determinants of efficacy remain largely unexplained, even after accounting for the presence of a provider-based component and a specified implementation strategy. Intermediate outcomes, or implementation outcomes (eg, acceptability, feasibility, and sustainability) [83,84], which are often not reported on in clinical research may help to further explain intervention efficacy. For instance, the degree to which patients and providers find particular technology-based alcohol interventions to be acceptable, feasible, and sustainable to use may influence the effectiveness of those interventions.

Future research should also strive to report more frequently on implementation strategies that occur during intervention studies [65]. Although outside the scope of this study, future reviews may benefit from examining whether or not the effectiveness of technology-based interventions can be predicted by the type of implementation strategy used. For instance, it is conceivable that variation in training and ongoing technical assistance, mandates from leadership, or efforts to engage patients and increase consumer demand for technology-based alcohol interventions could lead to greater effectiveness of these interventions. Relatedly, implementation studies should continue to examine the systematic use of strategies to support or improve the delivery of technology-based alcohol interventions.

Finally, this review highlights a potential lack of current integration between technology-based alcohol interventions and existing EHR systems; this technological integration has been posited to be a critical limiting factor in realizing the public health impact promised by technology-based behavioral health interventions [85]. Greater efforts to integrate technology-based alcohol interventions with existing EHR systems in primary care will be necessary to ensure the scale up and sustainability of technology-based alcohol interventions.

Limitations

We acknowledge that our systematic review may be limited to some degree by publication bias. It is important to recognize that there is a general bias, in both authors and publishers, toward prioritizing the publication of positive findings (ie, evidence in support of tested interventions) over null or negative findings. This can lead to misleadingly favorable conclusions about the efficacy of interventions. Although unable to fully address this limitation, we included dissertations in our systematic review to help mitigate this concern.

It is also worth noting that, given the nascent stage of the implementation science field, authors are often inconsistent in their reporting of implementation strategies used to deliver their interventions [65]. Some authors may not have reported or specified implementation strategies that were actually used, increasing the risk of reporting bias. Additionally, different authors may have specified the same implementation strategy in different ways. It is possible that these inconsistencies influenced our findings regarding the relationship between implementation strategies and efficacy.

As is generally the case in systematic reviews, our study is subject to the limitations of potential errors or biases in searching, including, and coding articles. For instance, it is possible that our search strategy failed to identify all relevant studies (eg, non-English language articles), that we were too conservative in our inclusion of relevant studies, or that we erroneously coded content in some articles. To offset these concerns, we (1) solicited the expertise of 2 university librarians to conduct extensive database and search string testing to ensure that our search strategy was appropriately comprehensive and (2) used 2 study investigators during the inclusion and coding processes to limit reviewer fatigue and ensure interrater agreement. Finally, we limited our review to primary care and general practice settings, and although previous reviews have addressed technology-based behavioral health interventions in the emergency department and social work settings [15,16,86], future research focused on other settings, including specialty care or dental health, may further contribute to this area of investigation.

Implications

This systematic review contributes substantially to the conceptualization of technology-based alcohol interventions, an understanding of the range of implementation contexts and formats in which these interventions are delivered, and an initial assessment of the efficacy and effect modifiers for technology-based alcohol interventions. The use of technology-based tools in primary care settings represents a promising approach to enhance the efficiency, service delivery flexibility, and effectiveness of interventions for alcohol-related problems. Our systematic review identifies that the past 2 decades have borne witness to the delivery of technology-based alcohol interventions through a variety of different technological platforms (eg, computer, Web, or mobile), settings (eg, clinic, home, or both), and targets (eg, patients, providers, or both). These findings provide initial support for the efficacy of technology-based alcohol interventions, particularly when deployed through a specific implementation strategy and involving a provider in the delivery process, and we encourage future research to further establish the efficacy, moderators of efficacy, and implementation strategies for delivering these types of interventions in primary care settings and beyond.

Acknowledgments

The authors would like to acknowledge Lori Siegel, Reference Librarian at the Washington University Brown School of Social Work and Laura Simon, Clinical Librarian at the Bernard Becker Medical Library, for their substantial assistance with this literature review. Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number K12DA041449, NIDA F31DA039776, NIMH T32MH019960, a grant from the Foundation for Barnes-Jewish Hospital, and the Washington University Institute for Public Health, Center for Dissemination and Implementation Pilot Program. Funding for this project was also provided by the Institute of Clinical and Translational Sciences through the Clinical and Translational Science Award Program of the National Center for Advancing Translational Sciences at the National Institutes of Health, Grant Number UL1 TR000448.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search string of 77 Boolean terms searched by Title, Abstract, and Subject.

[PDF File (Adobe PDF File), 68KB - jmir_v21i4e10859_app1.pdf]

Multimedia Appendix 2

Studies of technology-based alcohol interventions in primary care (n=42).

[PDF File (Adobe PDF File), 221KB - jmir_v21i4e10859_app2.pdf]

Multimedia Appendix 3

Summary of results on alcohol-related outcomes.

[PDF File (Adobe PDF File), 148KB - jmir_v21i4e10859_app3.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 Dec 22;392(10152):1015-1035. [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]
2. Hartz SM, Oehlert M, Horton AC, Grucza RA, Fisher SL, Culverhouse RC, et al. Daily drinking is associated with increased mortality. Alcohol Clin Exp Res 2018 Nov;42(11):2246-2255. [doi: 10.1111/acer.13886] [Medline: 30281161]
3. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of dsm-5 alcohol use disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiatry 2015 Aug;72(8):757-766 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.0584] [Medline: 26039070]
4. World Health Organization. 2014. Global status report on alcohol and health URL: https://www.who.int/substance_abuse/publications/global_alcohol_report/msb_gsr_2014_1.pdf?ua=1 [accessed 2019-03-04] [WebCite Cache ID 76dKbJ2L2]
5. May PA, Gossage JP, Kalberg WO, Robinson LK, Buckley D, Manning M, et al. Prevalence and epidemiologic characteristics of FASD from various research methods with an emphasis on recent in-school studies. Dev Disabil Res Rev 2009;15(3):176-192. [doi: 10.1002/ddrr.68] [Medline: 19731384]
6. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 national and state costs of excessive alcohol consumption. Am J Prev Med 2015 Nov;49(5):e73-e79. [doi: 10.1016/j.amepre.2015.05.031] [Medline: 26477807]
7. Kaner E, Bland M, Cassidy P, Coulton S, Dale V, Deluca P, et al. Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial. Br Med J 2013;346:e8501 [FREE Full text] [doi: 10.1136/bmj.e8501] [Medline: 23303891]
8. Pilowsky DJ, Wu L. Screening for alcohol and drug use disorders among adults in primary care: a review. Subst Abuse Rehabil 2012 Apr;3(1):25-34 [FREE Full text] [doi: 10.2147/SAR.S30057] [Medline: 22553426]
9. Glass JE, McKay JR, Gustafson DH, Kornfield R, Rathouz PJ, McTavish FM, et al. Treatment seeking as a mechanism of change in a randomized controlled trial of a mobile health intervention to support recovery from alcohol use disorders. J Subst Abuse Treat 2017 Jun;77:57-66. [doi: 10.1016/j.jsat.2017.03.011] [Medline: 28476273]
10. Wallace P, Bendtsen P. Internet applications for screening and brief interventions for alcohol in primary care settings - implementation and sustainability. Front Psychiatry 2014;5:151 [FREE Full text] [doi: 10.3389/fpsyt.2014.00151] [Medline: 25400593]
11. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares M, et al. Integrating addiction treatment into primary care using mobile health technology: protocol for an implementation research study. Implement Sci 2014 May 29;9:65 [FREE Full text] [doi: 10.1186/1748-5908-9-65] [Medline: 24884976]
12. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary care. Arch Intern Med 2008 Mar 10;168(5):530-536. [doi: 10.1001/archinternmed.2007.109] [Medline: 18332300]
13. Bertholet N, Daeppen J, Wietlisbach V, Fleming M, Burnand B. Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. Arch Intern Med 2005 May 9;165(9):986-995. [doi: 10.1001/archinte.165.9.986] [Medline: 15883236]
14. Harris SK, Louis-Jacques J, Knight JR. Screening and brief intervention for alcohol and other abuse. Adolesc Med State Art Rev 2014 Apr;25(1):126-156. [Medline: 25022191]
15. Ramsey AT, Montgomery K. Technology-based interventions in social work practice: a systematic review of mental health interventions. Soc Work Health Care 2014 Oct;53(9):883-899 [FREE Full text] [doi: 10.1080/00981389.2014.925531] [Medline: 25321935]
16. Choo EK, Ranney ML, Aggarwal N, Boudreaux ED. A systematic review of emergency department technology-based behavioral health interventions. Acad Emerg Med 2012 Mar;19(3):318-328. [doi: 10.1111/j.1553-2712.2012.01299.x] [Medline: 22435865]
17. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. J Med Internet Res 2014;16(6):e142 [FREE Full text] [doi: 10.2196/jmir.3193] [Medline: 24892426]
18. Kaner EF, Beyer FR, Muirhead C, Campbell F, Pienaar ED, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2018 Dec 24;2:1-248. [doi: 10.1002/14651858.CD004148.pub4] [Medline: 29476653]
19. 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 Dec 25;9:1-185. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
20. Nair NK, Newton NC, Shakeshaft A, Wallace P, Teesson M. A systematic review of digital and computer-based alcohol intervention programs in primary care. Curr Drug Abuse Rev 2015;8(2):111-118. [doi: 10.2174/1874473708666150916113538] [Medline: 26373848]
21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

22. 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 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
23. Bendtsen P, Stark Ekman D, Johansson A, Carlfjord S, Andersson A, Leijon M, et al. Referral to an electronic screening and brief alcohol intervention in primary health care in sweden: impact of staff referral to the computer. Int J Telemed Appl 2011:1-10 [FREE Full text] [doi: 10.1155/2011/918763] [Medline: 21603024]
24. Bischof G, Grothues JM, Reinhardt S, Meyer C, John U, Rumpf H. Evaluation of a telephone-based stepped care intervention for alcohol-related disorders: a randomized controlled trial. Drug Alcohol Depend 2008 Mar 1;93(3):244-251. [doi: 10.1016/j.drugalcdep.2007.10.003] [Medline: 18054443]
25. Brown RL, Saunders LA, Bobula JA, Mundt MP, Koch PE. Randomized-controlled trial of a telephone and mail intervention for alcohol use disorders: three-month drinking outcomes. Alcohol Clin Exp Res 2007 Aug;31(8):1372-1379. [doi: 10.1111/j.1530-0277.2007.00430.x] [Medline: 17550366]
26. Butler SF, Chiauzzi E, Bromberg JI, Budman SH, Buono DP. Computer-assisted screening and intervention for alcohol problems in primary care. J Technol Hum Serv 2003 Jun;21(3):1-19. [doi: 10.1300/J017v21n03_01]
27. Cucciare MA, Weingardt KR, Ghaus S, Boden MT, Frayne SM. A randomized controlled trial of a web-delivered brief alcohol intervention in Veterans Affairs primary care. J Stud Alcohol Drugs 2013 May;74(3):428-436. [doi: 10.15288/jsad.2013.74.428] [Medline: 23490572]
28. Dickinson WP, Glasgow RE, Fisher L, Dickinson LM, Christensen SM, Estabrooks PA, et al. Use of a website to accomplish health behavior change: if you build it, will they come? And will it work if they do? J Am Board Fam Med 2013;26(2):168-176 [FREE Full text] [doi: 10.3122/jabfm.2013.02.110344] [Medline: 23471930]
29. Dimeff L, McNeely M. Computer-enhanced primary care practitioner advice for high-risk college drinkers in a student primary health-care setting. Cogn Behav Pract 2000 Dec;7(1):82-100. [doi: 10.1016/S1077-7229(00)80010-3]
30. Fink A, Elliott MN, Tsai M, Beck JC. An evaluation of an intervention to assist primary care physicians in screening and educating older patients who use alcohol. J Am Geriatr Soc 2005 Nov;53(11):1937-1943. [doi: 10.1111/j.1532-5415.2005.00476.x] [Medline: 16274375]
31. Gryczynski J, Mitchell SG, Gonzales A, Moseley A, Peterson TR, Ondersma SJ, et al. A randomized trial of computerized vs in-person brief intervention for illicit drug use in primary care: outcomes through 12 months. J Subst Abuse Treat 2015 Mar;50:3-10 [FREE Full text] [doi: 10.1016/j.jsat.2014.09.002] [Medline: 25282578]
32. Harris SK, Csémy L, Sherritt L, Starostova O, van Hook S, Johnson J, et al. Computer-facilitated substance use screening and brief advice for teens in primary care: an international trial. Pediatrics 2012 Jun;129(6):1072-1082 [FREE Full text] [doi: 10.1542/peds.2011-1624] [Medline: 22566420]
33. Hasin DS, Aharonovich E, O'Leary A, Greenstein E, Pavlicova M, Arunajadai S, et al. Reducing heavy drinking in HIV primary care: a randomized trial of brief intervention, with and without technological enhancement. Addiction 2013 Jul;108(7):1230-1240 [FREE Full text] [doi: 10.1111/add.12127] [Medline: 23432593]
34. Helstrom AW, Ingram E, Wang W, Small D, Klaus J, Oslin D. Treating heavy drinking in primary care practices: evaluation of a telephone-based intervention program. Addict Disord Their Treat 2014;13(3):101-109. [doi: 10.1097/ADT.0b013e31827e206c]
35. Helzer JE, Rose GL, Badger GJ, Searles JS, Thomas CS, Lindberg SA, et al. Using interactive voice response to enhance brief alcohol intervention in primary care settings. J Stud Alcohol Drugs 2008 Mar;69(2):251-258. [doi: 10.15288/jsad.2008.69.251] [Medline: 18299766]
36. Holtrop JS, Dosh SA, Torres T, Thum YM. The community health educator referral liaison (CHERL): a primary care practice role for promoting healthy behaviors. Am J Prev Med 2008 Nov;35(5 Suppl):S365-S372. [doi: 10.1016/j.amepre.2008.08.012] [Medline: 18929983]
37. Kalapatapu RK, Ho J, Cai X, Vinogradov S, Batki SL, Mohr DC. Cognitive-behavioral therapy in depressed primary care patients with co-occuring problematic alcohol use: effect of telephone-administered vs face-to-face treatment-a secondary analysis. J Psychoactive Drugs 2014;46(2):85-92 [FREE Full text] [doi: 10.1080/02791072.2013.876521] [Medline: 25052784]
38. Kypri K, Saunders JB, Williams SM, McGee RO, Langley JD, Cashell-Smith ML, et al. Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. Addiction 2004 Nov;99(11):1410-1417. [doi: 10.1111/j.1360-0443.2004.00847.x] [Medline: 15500594]
39. Lin JC, Karno MP, Tang L, Barry KL, Blow FC, Davis JW, et al. Do health educator telephone calls reduce at-risk drinking among older adults in primary care? J Gen Intern Med 2010 Apr;25(4):334-339 [FREE Full text] [doi: 10.1007/s11606-009-1223-2] [Medline: 20101471]
40. McCausland C. A Randomized Clinical Trial of Computer Delivered Personal Feedback for Hazardous Drinkers in Primary Care. Unpublished Dissertation: The University of Memphis; 2011.
41. McMenamin J, Nicholson R, Leech K. Patient Dashboard: the use of a colour-coded computerised clinical reminder in Whanganui regional general practices. J Prim Health Care 2011 Dec 1;3(4):307-310. [doi: 10.1071/HC11307] [Medline: 22132384]

42. Moore AA, Blow FC, Hoffing M, Welgreen S, Davis JW, Lin JC, et al. Primary care-based intervention to reduce at-risk drinking in older adults: a randomized controlled trial. Addiction 2011 Jan;106(1):111-120 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03229.x] [Medline: 21143686]
43. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
44. Oslin DW, Sayers S, Ross J, Kane V, Ten HT, Conigliaro J, et al. Disease management for depression and at-risk drinking via telephone in an older population of veterans. Psychosom Med 2003;65(6):931-937. [doi: 10.1097/01.PSY.0000097335.35776.FB] [Medline: 14645769]
45. Pereira CA, Wen CL, Tavares H. Alcohol abuse management in primary care: an e-learning course. Telemed J E Health 2015 Mar;21(3):200-206. [doi: 10.1089/tmj.2014.0042] [Medline: 25599269]
46. Possemato K, Bishop TM, Willis MA, Lantinga LJ. Healthcare utilization and symptom variation among veterans using Behavioral Telehealth Center services. J Behav Health Serv Res 2013 Oct;40(4):416-426. [doi: 10.1007/s11414-013-9338-y] [Medline: 23616251]
47. Ruf D, Berner M, Kriston L, Lohmann M, Mundle G, Lorenz G, et al. Cluster-randomized controlled trial of dissemination strategies of an online quality improvement programme for alcohol-related disorders. Alcohol Alcohol 2010;45(1):70-78. [doi: 10.1093/alcalc/agp079] [Medline: 19889887]
48. Schwartz RP, Gryczynski J, Mitchell SG, Gonzales A, Moseley A, Peterson TR, et al. Computerized versus in-person brief intervention for drug misuse: a randomized clinical trial. Addiction 2014 Jul;109(7):1091-1098 [FREE Full text] [doi: 10.1111/add.12502] [Medline: 24520906]
49. Stoner SA, Mikko AT, Carpenter KM. Web-based training for primary care providers on screening, brief intervention, and referral to treatment (SBIRT) for alcohol, tobacco, and other drugs. J Subst Abuse Treat 2014;47(5):362-370 [FREE Full text] [doi: 10.1016/j.jsat.2014.06.009] [Medline: 25115136]
50. Tanner TB, Wilhelm SE, Rossie KM, Metcalf MP. Web-based SBIRT skills training for health professional students and primary care providers. Subst Abus 2012;33(3):316-320 [FREE Full text] [doi: 10.1080/08897077.2011.640151] [Medline: 22738013]
51. Vinson DC, Devera-Sales A. Computer‐generated written behavioral contracts with problem drinkers in primary medical care. Subst Abus 2000 Dec;21(4):215-222. [doi: 10.1080/08897070009511434] [Medline: 12466660]
52. Walton MA, Bohnert K, Resko S, Barry KL, Chermack ST, Zucker RA, et al. Computer and therapist based brief interventions among cannabis-using adolescents presenting to primary care: one year outcomes. Drug Alcohol Depend 2013 Oct 1;132(3):646-653 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.04.020] [Medline: 23711998]
53. Walton MA, Resko S, Barry KL, Chermack ST, Zucker RA, Zimmerman MA, et al. A randomized controlled trial testing the efficacy of a brief cannabis universal prevention program among adolescents in primary care. Addiction 2014 May;109(5):786-797 [FREE Full text] [doi: 10.1111/add.12469] [Medline: 24372937]
54. Williams EC, Achtmeyer CE, Kivlahan DR, Greenberg D, Merrill JO, Wickizer TM, et al. Evaluation of an electronic clinical reminder to facilitate brief alcohol-counseling interventions in primary care. J Stud Alcohol Drugs 2010 Sep;71(5):720-725 [FREE Full text] [Medline: 20731977]
55. Wongpakaran T, Petcharaj K, Wongpakaran N, Sombatmai S, Boriputtakul T, Intarakamhaeng D, et al. The effect of telephone-based intervention (TBI) in alcohol abusers: a pilot study. J Med Assoc Thai 2011 Jul;94(7):849-856. [Medline: 21774293]
56. Zanjani F, Bush H, Oslin D. Telephone-based psychiatric referral-care management intervention health outcomes. Telemed J E Health 2010 Jun;16(5):543-550 [FREE Full text] [doi: 10.1089/tmj.2009.0139] [Medline: 20575721]
57. Acosta MC, Possemato K, Maisto SA, Marsch LA, Barrie K, Lantinga L, et al. Web-delivered CBT reduces heavy drinking in OEF-OIF veterans in primary care with symptomatic substance use and PTSD. Behav Ther 2017 Dec;48(2):262-276 [FREE Full text] [doi: 10.1016/j.beth.2016.09.001] [Medline: 28270335]
58. Dawson-Rose C, Draughon JE, Cuca Y, Zepf R, Huang E, Cooper BA, et al. Changes in Specific Substance Involvement Scores among SBIRT recipients in an HIV primary care setting. Addict Sci Clin Pract 2017 Dec 12;12(1):34 [FREE Full text] [doi: 10.1186/s13722-017-0101-1] [Medline: 29229000]
59. Duru OK, Xu H, Moore AA, Mirkin M, Ang A, Tallen L, et al. Examining the impact of separate components of a multicomponent intervention designed to reduce at risk drinking among older adults: the Project SHARE study. Alcohol Clin Exp Res 2015 Jul;39(7):1227-1235 [FREE Full text] [doi: 10.1111/acer.12754] [Medline: 26033430]
60. Hunter R, Wallace P, Struzzo P, Vedova RD, Scafuri F, Tersar C, et al. Randomised controlled non-inferiority trial of primary care-based facilitated access to an alcohol reduction website: cost-effectiveness analysis. BMJ Open 2017 Nov 3;7(11):e014577 [FREE Full text] [doi: 10.1136/bmjopen-2016-014577] [Medline: 29102983]
61. Knight JR, Kuzubova K, Csemy L, Sherritt L, Copelas S, Harris SK. Computer-facilitated screening and brief advice to reduce adolescents' heavy episodic drinking: a study in two countries. J Adolesc Health 2018 Jan;62(1):118-120 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.08.013] [Medline: 29054734]

62. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
63. Rose GL, Badger GJ, Skelly JM, MacLean CD, Ferraro TA, Helzer JE. A randomized controlled trial of brief intervention by interactive voice response. Alcohol Alcohol 2017 May 1;52(3):335-343 [FREE Full text] [doi: 10.1093/alcalc/agw102] [Medline: 28069598]
64. Grol R, Wensing M, Eccles M, Davis D. Improving Patient Care: The Implementation of Change in Health Care. Chichester: John Wiley & Sons Inc; 2013.
65. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 1;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
66. Improving the Quality of Health Care for Mental and Substance-Use Conditions: Quality Chasm Series. Washington, DC: National Academies Press; 2006.
67. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci 2009;4:18 [FREE Full text] [doi: 10.1186/1748-5908-4-18] [Medline: 19351400]
68. Ramsey A, Lord S, Torrey J, Marsch L, Lardiere M. Paving the way to successful implementation: identifying key barriers to use of technology-based therapeutic tools for behavioral health care. J Behav Health Serv Res 2016 Jan;43(1):54-70. [doi: 10.1007/s11414-014-9436-5] [Medline: 25192755]
69. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile application in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
70. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr;69(2):123-157 [FREE Full text] [doi: 10.1177/1077558711430690] [Medline: 22203646]
71. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
72. Powell BJ, Proctor EK, Glass JE. A systematic review of strategies for implementing empirically supported mental health interventions. Res Soc Work Pract 2014 Mar 1;24(2):192-212 [FREE Full text] [doi: 10.1177/1049731513505778] [Medline: 24791131]
73. Ryan R, Hill S, Prictor M, McKenzie J. Cochrane Consumers and Communication Review Group. 2013 May. Study Quality Guide URL: https://cccrg.cochrane.org/author-resources [accessed 2017-11-27] [WebCite Cache ID 76dV863ZN]
74. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Br Med J 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
75. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):178-186. [doi: 10.1016/j.cpr.2010.10.002] [Medline: 21095051]
76. Gainsbury S, Blaszczynski A. A systematic review of Internet-based therapy for the treatment of addictions. Clin Psychol Rev 2011 Apr;31(3):490-498. [doi: 10.1016/j.cpr.2010.11.007] [Medline: 21146272]
77. Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to improve the selection and tailoring of implementation strategies. J Behav Health Serv Res 2017 Apr;44(2):177-194 [FREE Full text] [doi: 10.1007/s11414-015-9475-6] [Medline: 26289563]
78. Lyon AR, Lewis CC. Designing health information technologies for uptake: development and implementation of measurement feedback systems in mental health service delivery. Adm Policy Ment Health 2016 May;43(3):344-349 [FREE Full text] [doi: 10.1007/s10488-015-0704-3] [Medline: 26658691]
79. Ornstein SM, Miller PM, Wessell AM, Jenkins RG, Nemeth LS, Nietert PJ. Integration and sustainability of alcohol screening, brief intervention, and pharmacotherapy in primary care settings. J Stud Alcohol Drugs 2013 Jul;74(4):598-604 [FREE Full text] [Medline: 23739024]
80. Williams EC, Johnson ML, Lapham GT, Caldeiro RM, Chew L, Fletcher GS, et al. Strategies to implement alcohol screening and brief intervention in primary care settings: a structured literature review. Psychol Addict Behav 2011 Jun;25(2):206-214. [doi: 10.1037/a0022102] [Medline: 21517141]
81. Glisson C, Schoenwald SK, Hemmelgarn A, Green P, Dukes D, Armstrong KS, et al. Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. J Consult Clin Psychol 2010 Aug;78(4):537-550 [FREE Full text] [doi: 10.1037/a0019160] [Medline: 20658810]
82. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
83. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]

84. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
85. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
86. Haskins BL, Davis-Martin R, Abar B, Baumann BM, Harralson T, Boudreaux ED. Health evaluation and referral assistant: a randomized controlled trial of a web-based screening, brief intervention, and referral to treatment system to reduce risky alcohol use among emergency department patients. J Med Internet Res 2017 May 1;19(5):e119 [FREE Full text] [doi: 10.2196/jmir.6812] [Medline: 28461283]

Abbreviations

ANOVA: analysis of variance
EBI: evidence-based intervention
EHR: electronic health record
TAU: treatment as usual

Edited by G Eysenbach; submitted 23.04.18; peer-reviewed by J Glass, K Champion, A Fink; comments to author 03.09.18; revised version received 29.01.19; accepted 30.01.19; published 08.04.19.

Please cite as:
Ramsey AT, Satterfield JM, Gerke DR, Proctor EK
Technology-Based Alcohol Interventions in Primary Care: Systematic Review
J Med Internet Res 2019;21(4):e10859
URL: https://www.jmir.org/2019/4/e10859/
doi:10.2196/10859
PMID:30958270

Original Paper

Addressing Diabetes and Poorly Controlled Hypertension: Pragmatic mHealth Self-Management Intervention

Allison A Lewinski[1*], MPH, PhD; Uptal D Patel[2,3,4*], MD; Clarissa J Diamantidis[2,5,6*], MHS, MD; Megan Oakes[6*], MS; Khaula Baloch[7*], MPH; Matthew J Crowley[1,8*], MHS, MD; Jonathan Wilson[9*], MS; Jane Pendergast[9*], PhD; Holly Biola[10*], MPH, MD; L Ebony Boulware[5*], MPH, MD; Hayden B Bosworth[1,5,6,11,12*], PhD

[1]Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, United States

[2]Division of Nephrology, Department of Medicine, Duke University School of Medicine, Durham, NC, United States

[3]Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, United States

[4]Gilead Sciences, Inc, Foster City, CA, United States

[5]Division of General Internal Medicine, Department of Medicine, Duke University School of Medicine, Durham, NC, United States

[6]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, United States

[7]Outcomes Research Group, Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, United States

[8]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Duke University School of Medicine, Durham, NC, United States

[9]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, United States

[10]Department of Family Medicine, Lincoln Community Health Center, Durham, NC, United States

[11]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States

[12]School of Nursing, Duke University, Durham, NC, United States

[*]all authors contributed equally

Corresponding Author:
Hayden B Bosworth, PhD
Durham Center of Innovation to Accelerate Discovery and Practice Transformation
Durham Veterans Affairs Health Care System
Attn: HSR&D COIN (558/152)
508 Fulton St
Durham, NC,
United States
Phone: 1 919 286 0411 ext 7101
Email: hayden.bosworth@duke.edu

Abstract

Background: Patients with diabetes and poorly controlled hypertension are at increased risk for adverse renal and cardiovascular outcomes. Identifying these patients early and addressing modifiable risk factors is central to delaying renal complications such as diabetic kidney disease. Mobile health (mHealth), a relatively inexpensive and easily scalable technology, can facilitate patient-centered care and promote engagement in self-management, particularly for patients of lower socioeconomic status. Thus, mHealth may be a cost-effective way to deliver self-management education and support.

Objective: This feasibility study aimed to build a population management program by identifying patients with diabetes and poorly controlled hypertension who were at risk for adverse renal outcomes and evaluate a multifactorial intervention to address medication self-management. We recruited patients from a federally qualified health center (FQHC) in an underserved, diverse county in the southeastern United States.

Methods: Patients were identified via electronic health record. Inclusion criteria were age between 18 and 75 years, diagnosis of type 2 diabetes, poorly controlled hypertension over the last 12 months (mean clinic systolic blood pressure [SBP] ≥140 mm Hg and/or diastolic blood pressure [DBP] ≥90 mm Hg), access to a mobile phone, and ability to receive text messages and emails. The intervention consisted of monthly telephone calls for 6 months by a case manager and weekly, one-way informational text messages. Engagement was defined as the number of phone calls completed during the intervention; individuals who completed 4 or more calls were considered engaged. The primary outcome was change in SBP at the conclusion of the intervention.

Results: Of the 141 patients enrolled, 84.0% (118/141) of patients completed 1 or more phone calls and had follow-up SBP measurements for analysis. These patients were on average 56.9 years of age, predominately female (73/118, 61.9%), and nonwhite by self-report (103/118, 87.3%). The proportion of participants with poor baseline SBP control (50/118, 42.4%) did not change significantly at study completion (53/118, 44.9%) (*P*=.64). Participants who completed 4 or more phone calls (98/118, 83.1%) did not experience a statistically significant decrease in SBP when compared to those who completed fewer calls.

Conclusion: We did not reduce uncontrolled hypertension even among the more highly engaged. However, 83% of a predominately minority and low-income population completed at least 67% of the multimodal mHealth intervention. Findings suggest that combining an automated electronic health record system to identify at-risk patients with a tailored mHealth protocol can provide education to this population. While this intervention was insufficient to effect behavioral change resulting in better hypertension control, it does suggest that this FQHC population will engage in low-cost population health applications with a potentially promising impact.

Trial Registration: ClinicalTrials.gov NCT02418091; https://clinicaltrials.gov/ct2/show/NCT02418091 (Archived by WebCite at http://www.webcitation.org/76RBvacVU)

(J Med Internet Res 2019;21(4):e12541) doi:10.2196/12541

KEYWORDS

telemedicine; cardiovascular diseases; diabetes mellitus type 2; vulnerable populations; renal insufficiency; professional-patient relations; hypertension

Introduction

Diabetes and hypertension are two of the most prevalent chronic illnesses worldwide. Patients with comorbid diabetes and hypertension are at greater risk for progressive renal and cardiovascular complications, including diabetic kidney disease (DKD) [1-6]. Improving long-term outcomes in this high-risk group depends on identifying patients with diabetes and poorly controlled hypertension, addressing modifiable risk factors, and ensuring access to optimal treatment and education early in the patient's disease trajectory [4,7].

The incidence and prevalence of diabetes and hypertension continue to disproportionately impact individuals who are minorities and/or of lower socioeconomic status populations [7-9]. Reasons for higher rates of diabetes and hypertension in these populations can include limited access to quality medical care, low health literacy, and lack of insurance [7,10,11]. These challenges and barriers to engaging in self-management place patients of lower socioeconomic status populations at risk for poorer health and adverse renal and cardiovascular outcomes [12]. Identifying methods to deliver personalized, disease-specific support may help decrease the potential impact of diabetes and poorly controlled hypertension.

Patients with diabetes and poorly controlled hypertension benefit from aggressive treatment that improves control of modifiable risk factors such as management of blood glucose, blood pressure, diet, exercise, and weight and smoking cessation [13-15]. Simultaneously addressing these factors may preserve renal function and delay a decline in renal function and death [7-9]. Patient-centered interventions that address multiple factors and facilitate personalized problem solving may be more effective than interventions with a generalized approach [3].

Mobile health (mHealth) is one way to provide patient-centered information and promote engagement in self-management for patients who are at high risk for DKD. Mobile health interventions are effective ways to change behaviors in individuals with chronic illnesses such as diabetes and hypertension [16-18]. Interventions via mHealth have high potential reach, as a large percentage of US adults have a mobile phone [19], including individuals with low socioeconomic status [20-22]. Additionally, mHealth technologies are inexpensive and can be easily scaled, thus increasing the potential for behavioral intervention dissemination. However, there has been a lack of studies using mHealth to reduce risk factors among ethnic and racial minorities and low-income individuals who are at particularly high risk for poor renal and cardiovascular outcomes [23,24].

The purpose of this feasibility study was to pilot test an mHealth intervention among individuals with diabetes and poorly controlled hypertension. The intervention included tailored behavioral-educational components with a focus on disease self-management.

Methods

Study Design

This was a single-arm, pragmatic study designed to implement a 6-month intervention. This multifactorial intervention simultaneously addressed multiple risk factors for adverse renal outcomes through a combination of patient self-monitoring, behavioral therapies, and education that optimized adherence and improvements in health behavior self-efficacy. As a proxy for improvements in these process measures, we examined whether the intervention was associated with a positive impact on systolic blood pressure. Duke University's institutional review board approved this study (Pro00052081), and this study was registered with ClinicalTrials.gov (NCT02418091).

Setting

We recruited patients from a federally qualified health center (FQHC) in a midsize city in the southeastern United States. We chose this FQHC due to the high incidence of poorly controlled diabetes and hypertension in this underserved population. In 2017, this FQHC provided comprehensive primary and preventive care to over 33,500 unique individuals. This FQHC serves a population that is predominantly Latinx (47%) or black (39%), and many patients in this clinic system (50% of adults and 24% of children) are uninsured. Overall, patients seeking care in this clinic are lower income: 38% of patient households make less than 100% of the federal poverty level and only 19% make above 100% of the federal poverty level (43% did not provided income documentation).

Recruitment and Enrollment

We identified eligible patients using the FQHC’s electronic health record (EHR). Inclusion criteria for this study included age between 18 and 75 years, diagnosis of type 2 diabetes (International Classification of Disease [ICD]-9 codes 250.x0, 250.x2; ICD-10 codes E11.0-E11.9), poorly controlled hypertension (1-year mean clinic systolic blood pressure [SBP] ≥140 mm Hg and/or diastolic blood pressure [DBP] ≥90 mm Hg), and access to a mobile phone and ability to receive text messages and emails. Exclusion criteria included inability to speak English, residence in a nursing home or long-term care facility or receipt of home health care, or a current diagnosis of pancreatic insufficiency or diabetes secondary to pancreatitis. We excluded patients who self-reported alcohol use of more than 14 alcoholic beverages per week because this study was not designed to address substance abuse behaviors.

We identified patients using the FQHC’s EHR and then screened and identified eligible patients. Eligible patients received a letter from their primary care provider requesting study participation. After a 10-day period in which potential patients could opt out of the study, patients were contacted by the call center affiliated with the local academic clinical research organization. Patients were informed and consented over the phone using an institutional review board–approved script. All patients who verbally consented to the study were enrolled.

Simultaneous Risk Factor Control Using Telehealth to Slow Progression of Diabetic Kidney Disease Automated Population Program

The Simultaneous Risk Factor Control Using Telehealth to Slow Progression of Diabetic Kidney Disease Automated Population Program (STOP-DKD APP) consisted of two novel platforms. The first platform was an electronic registry that used the clinic’s EHR to identify the target population (eg, patients with diabetes and poorly controlled hypertension). This was combined with a second electronic platform that delivered an evidence-based behavioral intervention to improve self-management comprising educational content on diabetes and hypertension [25]. The goal of the intervention was not to replace clinic-based management but to supplement it in order to more efficiently intensify therapy that may otherwise be left until future appointments. Information for all eligible patients was entered into the STOP-DKD APP platform. The STOP-DKD APP intervention was guided by three behavioral science models: chronic care model [26,27], health decision model [28], and the transtheoretical model [29]. These complementary models informed the design of the intervention to slow DKD progression in the population.

The chronic care model describes factors that can improve functional and clinical outcomes, particularly in chronic conditions [26,27]. Patients with DKD and uncontrolled hypertension often receive suboptimal care due to fragmented and poorly designed health systems. Thus, the model acknowledges that a substantial portion of chronic care takes place outside of formal health care settings and highlights six core elements for the provision of optimal care of patients with chronic disease. The STOP-DKD APP intervention addressed the core elements described in the chronic care model [26,27] in order to create a more proactive provider team and more engaged patient by (1) engaging health systems interested in improvement strategies, (2) leveraging an innovative clinical information system to identify study patients and community resources to tailor information and feedback, (3) using decision support informed by health behavior models to optimize patient self-management, and (4) redesigning the delivery system by using case managers to facilitate management of complex medication regimens in close communication with participant primary care providers [30-32]. These intervention strategies created a more proactive provider team and more active patient.

We used the health decision model to guide the selection of behaviors related to treatment adherence, diabetes, and poorly controlled hypertension [28]. For patients with DKD and poorly controlled hypertension, the complexity of care requires that appropriate behavior change theories be applied toward understanding behaviors related to treatment adherence. To focus on health decisions, the health decision model draws upon other behavioral models to combine the influences of health beliefs and modifying factors with contributions from the patient preference literature, including important factors such as memory and the experience of side effects associated with medications [28,29,33,34]. The health decision model also identifies potential behavioral factors that may explain poor disease control related to treatment adherence by examining factors that hinder or promote health behaviors.

Behavior change theories are also used for understanding behaviors related to treatment adherence. Understanding the factors that hinder or promote health behaviors is central to the transtheoretical model [29], which we used to guide the incorporation of patient-centered, tailored information and feedback as the intervention focused heavily on the initiation and maintenance phases of behavior change [29]. While a generic health care–administered intervention may improve treatment adherence through reminders, a tailored intervention can address issues that are specifically relevant to each patient.

Thus, drawing on the stages of change [29] and the revised health decision model [28,29], the STOP DKD APP intervention helped patients to (1) set healthy goals and gain self-efficacy, (2) implement healthy behaviors and monitor performance, and (3) maintain the behaviors and associated risk factor control over time [35,36]. Together, these models directed the structure, content, and format of the intervention in order to optimize medical management of this population while encouraging patient engagement in and adherence to their self-management behaviors.

The telephone-based intervention was administered by a call center using nonclinician case managers. The intervention consisted of (1) monthly telephone calls for 6 months, (2) monthly emails that summarized the content covered with the case manager during the monthly phone call, and (3) weekly text messages. Table 1 provides examples of the content in the text messages and monthly calls. Patients received medication adherence information and e-reminders to take DKD-related medications and self-monitor their health status (eg, SBP, weight). This study did not include medical management for hypertension. All intervention components were designed to be culturally sensitive, and content addressed self-management facilitators and barriers common in this population. Additionally, all content provided was tailored to patient responses about self-management behaviors, medication adherence, smoking status, and prescribed medications. Patient responses were obtained during interactions with the case manager or the interactive text messages received during the intervention. All content was presented at a literacy level below a 6th grade reading level.

The monthly call was a review of education and self-management topics related to diabetes, hypertension, and hyperlipidemia. During these calls, the case manager reviewed the patient's current medications, changes in medication status, or side effects the patient was experiencing and then provided self-management content based on the patient's self-reported health status. Phone calls provided patients the opportunity to obtain personalized, disease-specific information and feedback. If a patient could not be reached on schedule, the case manager attempted to complete the phone call at the next scheduled call time. Once a call was completed, the call schedule was adjusted accordingly. Following the completion of each monthly phone call, patients were sent an email that summarized the content discussed in the phone call.

Patients received text messages containing information related to self-management and medication adherence on diabetes and hypertension, including messages on side effects, risks and benefits of medication treatment, and barriers to medication adherence. These text messages prompted patients to engage in self-management behaviors ("Remember to carry a snack or a source of sugar with you in case your blood sugar gets low"), provided education ("Exercise lowers your risk for heart disease and stroke, relieves stress, and strengthens your heart, muscles, and bones"), and offered suggestions ("You can decrease your salt intake by cutting back on fast foods and processed foods such as canned soups and vegetables and frozen dinners" and "To help prevent low blood sugar, eat your meals and snacks at the same time each day. Do not skip meals"). Text messages were sent 3 times each week around 6 pm for 6 months.

Measures

The focus of the intervention was to optimize medication management related to treatment for diabetes and hypertension. The primary study outcome was change in SBP, operationalized as controlled (SBP <140 mm Hg) or poorly controlled (SBP ≥140 mm Hg) from baseline to 90 days after the last completed phone call. Baseline SBP was defined as the SBP closest to study enrollment within a window from 1 year prior to 14 days into the study period. Completion SBP was defined as the closest SBP measurement to 90 days after the last completed phone call from a window spanning last call day to 180 days later. Our selection of these intervals represents our pragmatic approach to address potential sparsity of data for some individuals. Engagement was recorded as the number of completed monthly calls at the time of each SBP measurement. Individuals who completed 4 or more calls were considered engaged during the study period for pre-post comparisons, while those completing fewer were considered nonengaged.

Analyses

We examined engagement in the intervention to determine feasibility of providing an mHealth intervention to patients from an FQHC. We defined level of engagement as the number of completed phone calls during the intervention.

We used two statistical methods to evaluate the effectiveness of the STOP-DKD APP intervention. First, we used the McNemar test [37,38] to compare the proportion of patients who went from having a controlled blood pressure (SBP <140 mm Hg) or poorly controlled pressure (SBP ≥140 mm Hg) at baseline to the opposite category at the end of the study period. Second, using linear regression, we compared change in SBP over the time-in-study of those patients who had a controlled SBP at baseline to those patients who had a high SBP at baseline, adjusting for the level of study engagement and its interaction with baseline SBP. This approach allowed testing for a change over time, whether greater participation in the intervention was associated with a greater impact on SBP, and if that impact was different for those with high versus controlled SBP at baseline.

Table 1. Description of case manager call and text message topics.

Month	Phone calls	Text messages
1	Review of medications and side effects	Weekly medication reminders and study introduction
	Hypertension and cardiovascular disease knowledge	Stress and increased blood pressure
2	Review of medications and side effects	Weekly medication reminders
	Hypoglycemia and foot care	Hyperglycemia recognition
	Diabetes medications and side effects	Tobacco use
3	Review of medications and side effects	Weekly medication reminders
	Hyperlipidemia medications and side effects	Alcohol
	Diet and weight	Sleep health
	Hyperlipidemia knowledge	Cholesterol and blood pressure knowledge
4	Review of medications and side effects	Weekly medication reminders
	Hypertension medications	Patient and provider or clinic communication
	Exercise	Blood pressure and diabetes knowledge
	Sleep	Diet (lowering carbohydrates)
5	Review of medications and side effects	Weekly medication reminders
	Depression	Diet (carbohydrates and fiber)
	Tobacco use	Cholesterol knowledge
6	Review of medications, side effects, and aspirin use	Weekly medication reminders
	Alcohol knowledge	Blood pressure and blood sugar goals
	Patient-provider interaction	Patient and provider or clinic communication

Results

Sample Characteristics

We contacted 379 patients about the STOP DKD APP study; of those patients, we excluded 238. We enrolled 141 patients from May 2015 through January 2016 (see Figure 1 for the Consolidated Standards of Reporting Trials diagram). In total, 127 patients completed at least one phone call. Of those 127 patients, 125 had a baseline SBP and 118 had both a baseline and follow-up SBP. Therefore, the following analyses focused on the 118 patients (118/141, 83.7%) who completed 1 or more phone call and had both a baseline and follow-up SBP. These analyses do not include patients who withdrew from the study (6/141, 4.3%), were lost to follow-up (4/141, 2.8%), did not complete any phone calls (4/141, 2.8%), or did not have an eligible baseline or follow-up SBP (9/141, 6.4%).

Among the 118 individuals, the mean age was 56.9 years. Patients were primarily female (73/118, 61.9%), self-identified as nonwhite or black (103/118, 87.3%), and had a high school education or more (91/118, 77.1%). Most had no insurance (35/118, 29.7%) or government-funded support or special programs (53/118, 44.9%), while 25.4% (30/118) had private/commercial insurance. In regard to baseline clinical characteristics, the mean 12-month prior SBP of the sample was 139.5 (SD 19.8) mm Hg and mean DBP was 82.5 (SD 11.2) mm Hg and most had an estimated glomerular filtration rate (eGFR) >60 (74/118, 62.7%). The sample is fully described in Table 2.

Figure 1. Consolidated Standards of Reporting Trials diagram.

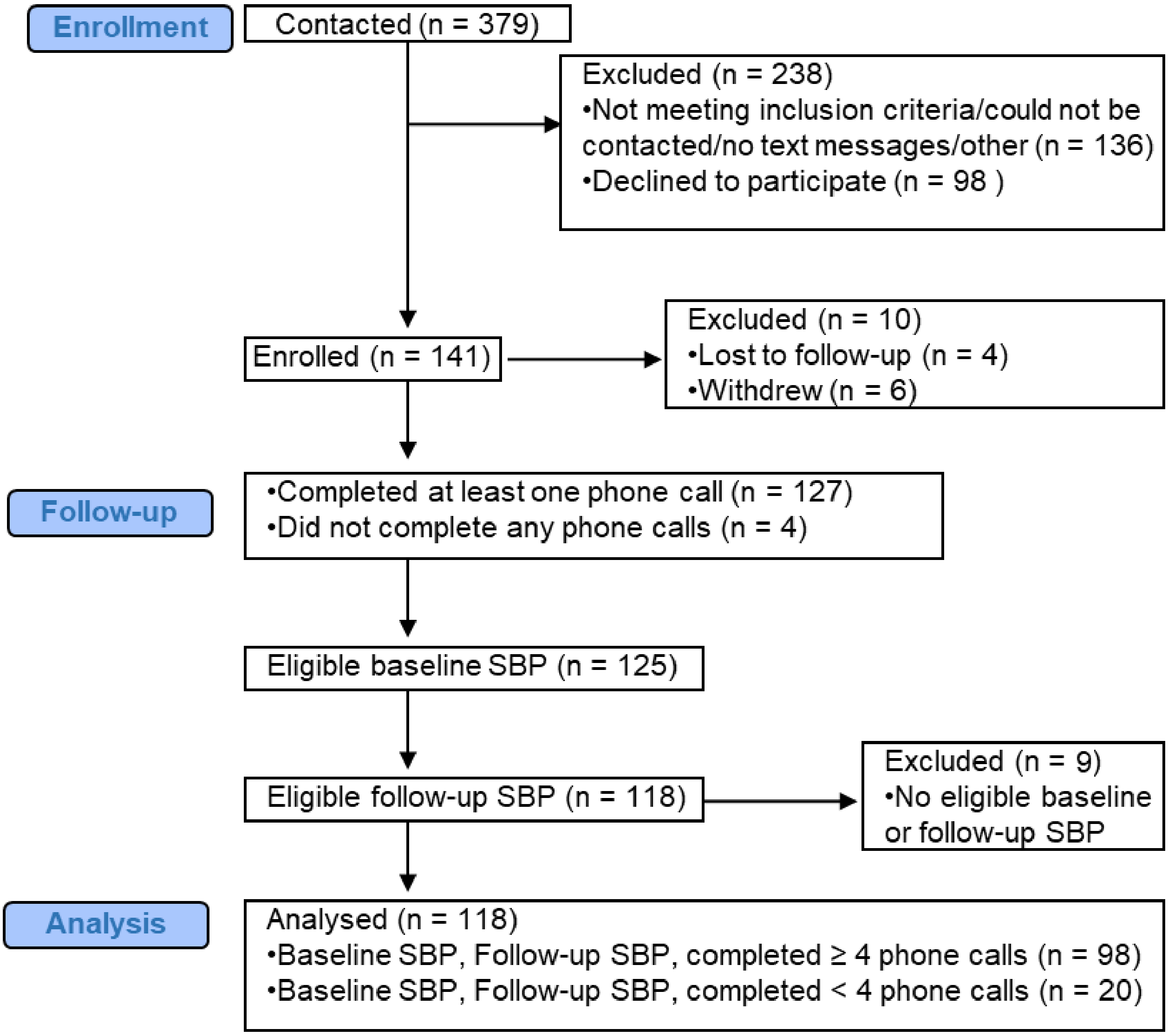


Table 2. Baseline demographic and clinical characteristics (n=118).

Characteristic	Value
Age in years, mean (SD)	56.9 (7.6)
Gender, female, n (%)	73 (61.9)
Race, n (%)	
White	15 (12.7)
Black	99 (83.9)
Other	4 (3.4)
Ethnicity (Hispanic) , n (%)	
Yes	2 (1.7)
No	114 (96.6)
Refused	2 (1.7)
Marital status, married, n (%)	25 (21.2)
Missed medication, yes, n (%)	57 (48.3)
Insurance status, n (%)	
Self-pay	35 (29.7)
Commercial	30 (25.4)
Government or special program	53 (44.9)
Access to computer, n (%)	
Yes	54 (45.8)
No	63 (53.4)
Missing	1 (0.8)
Education, high school graduate or greater, n (%)	91 (77.1)
Ability to do usual activities, no problems reported, n (%)	65 (55.1)
Ability to take care of self, no problems washing or dressing, n (%)	101 (85.6)
Self-rated health at baseline (0 to 100), mean (SD)	65.6 (23.6)
Systolic blood pressure (mm Hg), mean (SD)	139.5 (19.8)
Diastolic blood pressure (mm Hg), mean (SD)	82.5 (11.2)
Estimated glomerular filtration rate, n (%)	
<60	24 (20.3)
>60	74 (62.7)
Missing	20 (16.9)

Table 3. Baseline demographic and clinical characteristics by engagement.

Characteristic	Nonengaged (n=20)	Engaged (n=98)
Age in years, mean (SD)	50.3 (8.8)	58.2 (6.6)
Gender, female, n (%)	16 (80)	57 (58)
Race, n (%)		
White	3 (15)	12 (12)
Black	17 (85)	82 (84)
Other	0 (0)	4 (4)
Ethnicity (Hispanic), n (%)		
Yes	0 (0)	2 (2)
No	20 (100)	94 (96)
Refused	0 (0)	2 (2)
Marital status, married, n (%)	5 (25)	20 (20)
Missed medication, yes, n (%)	10 (50)	47 (48)
Insurance status, n (%)		
Self-pay	5 (25)	30 (31)
Commercial	4 (20)	26 (27)
Government or special program	11 (55)	42 (43)
Access to computer, n (%)		
Yes	9 (45)	45 (46)
No	11 (55)	52 (53)
Missing	0 (0)	1 (1)
Education, high school graduate or greater, n (%)	17 (85)	74 (76)
Ability to do usual activities, no problems reported, n (%)	8 (40)	57 (58)
Ability to take care of self, no problems washing or dressing, n (%)	18 (90)	83 (85)
Self-rated health at baseline (0-100), mean (SD)	70.1 (23.7)	64.6 (23.5)
Systolic blood pressure (mm Hg), mean (SD)	138.2 (21.9)	139.8 (19.5)
Diastolic blood pressure (mm Hg), mean (SD)	85.1 (7.6)	82.0 (11.8)
Estimated glomerular filtration rate, n (%)		
<60	3 (15)	21 (21)
>60	13 (65)	61 (62)
Missing	4 (20)	16 (16)

Engagement

We noted a few differences between engaged and nonengaged patients. Engaged patients (completed 4 or more phone calls) had a significantly higher mean age (*P*=.001, Wilcoxon rank-sum test) and were marginally more likely to be male (*P*=.08, Fisher exact test) and have a diagnosis of hypertension than nonengaged patients (*P*=.09, Fisher exact test). The baseline demographic and clinical characteristics by study engagement over the course of the intervention are presented in Table 3, and other than those reported above, no other differences were seen between engaged and nonengaged participants.

Table 4. Pre-post change in systolic blood pressure control status.

Baseline SBP	Postintervention SBP[a]		
	Control	Poor control	Total
Control	46 (39.0)	22 (18.6)	68
Poor control	19 (16.1)	31 (26.3)	50
Total	65	53	118

[a] SBP: systolic blood pressure.

Clinical Outcomes

The proportion of patients with poorly controlled SBP at the start of the study (50/118, 42.4%) did not change significantly at study completion (53/118, 44.9%) (*P*=.64). The proportions who switched from in control (defined as SBP <140 mm Hg) to poor control (defined as SBP ≥140 mm Hg) (22/118, 18.6%) and poor control to in control (19/118, 16.1%) were comparable (*P*=.64, McNemar test) (Table 4).

The McNemar test does not consider the magnitude of changes over time, and slight changes could reclassify a patient into the other group. Thus, we compared the rates of change in SBP while accounting for baseline SBP and engagement with the study. Participants who completed 4 or more phone calls (98/118, 83.1%) did not experience a statistically significant decrease in SBP when compared to those who completed fewer calls.

Discussion

Principal Findings

This study is among the first to examine the use of mHealth supported with phone calls from nonclinician case managers in a predominately minority and low-income sample. We successfully recruited and engaged a sample of patients with diabetes and poorly controlled hypertension from an FQHC for a multifactorial behavioral-educational population intervention primarily using the EHR. Individuals who completed 4 or more phone calls (out of 6) did not experience a statistical decrease in SBP relative to those patients who were not as engaged. Overall, the findings provide insight into designing population health management programs that aim to address modifiable risk factors for patients with both diabetes and poorly controlled hypertension.

The identification of patients using a system linked to an EHR can quickly detect patients at high risk for poor diabetes outcomes. EHRs can facilitate population health management because they facilitate the collection of patient data on a large scale and enable a more rapid and efficient analysis of these patient data [39]. The early identification of patients at risk for poor health outcomes, such as those with diabetes and poorly controlled hypertension, can decrease the progression of renal and cardiovascular complications and potentially decrease disparities in treatment and care [40]. The STOP-DKD APP study successfully facilitated the identification of patients at risk for DKD via a community clinic's EHR and provided education tailored to each patient.

Our study successfully recruited a black sample with limited financial resources from an FQHC for an mHealth intervention. The sample was highly engaged, with the majority completing 4 or more case manager–administered, telephone-delivered self-management education calls. One potential reason for high engagement could be the bidirectional monthly interaction with a case manager. These frequent encounters with the case manager may have addressed the patient's current questions about self-management and medication adherence in a more timely fashion than an episodic appointment with a medical provider [41,42]. This study's pragmatic design was an additional strength as there was no in-person contact, the intervention was administered by telephone, and blood pressure was assessed via each patient's EHR. Taken together, these strengths indicate that mHealth interventions in which routine encounters are delivered by nonclinicians can be used with high-risk, low-income populations without placing undue burden on clinicians at community health clinics.

There are several potential explanations for why this multifactorial intervention was not as impactful as hypothesized. First, our sample's blood pressure was relatively well controlled at baseline, which may have limited our ability to detect the clinical impact of improved SBP. Second, the intervention's dose of self-management education and/or the length of the intervention may not have been sufficient to affect cardiovascular disease risk factor control in this low-income population. Additionally, clinic staff may have given more attention to patients at baseline who had poorly controlled hypertension and may have focused less on those patients who had controlled blood pressure at baseline. Collectively, our findings reflect the literature that further research is needed on pragmatic, multifactorial interventions that address chronic illness self-management [43]. A greater understanding of this population of low-income individuals with chronic illness will help identify the optimal intervention dose and length, intervention strategies, and message content to impact metabolic outcomes.

Another reason for the limited impact on blood pressure may have been due to the use of routine clinic measurements for outcome ascertainment. The variance in frequency of blood pressure values for each patient may have impacted our results due to the imprecise nature of voluntary patient visits to the FQHC. For example, to obtain the SBP and DBP from the FQHC's EHR, we aimed to get as close as possible to a 3-month follow-up window during the 6-month study period. As a result, this limited our ability to collect data that was in sync with the completed phone calls. The SBP measurements used in this study for each individual were a result of the patients voluntarily going to the FQHC and having a blood pressure value logged into the system during the intervention time period. Additionally, the precision of the blood pressure readings may have been low, as only one blood pressure reading may have been taken during each clinic visit [44]. Variability in routine clinic blood pressure measurement has been increasingly recognized [45,46], so although university-affiliated clinics have standard procedures for BP measurement, it is possible that routine clinic practice and all available clinic data may have introduced error into our analysis. Our experience highlights the challenges of using clinical measures for outcomes and indicates the need for further research on pragmatic methods to accurately obtain clinical outcome data from patients who receive care at an FQHC.

Limitations

There are several limitations to this study. First, we sampled individuals who received medical care at an FQHC, which may limit generalizability of these results. Second, the patients in this study had a mean SBP of 140 mm Hg at baseline. The wide variance in the mean SBP may be because we included patients with any SBP over 140 mm Hg, which may have led to a mean SBP at baseline near the target threshold. Third, potential reasons for the minimal results seen with this study could be because of limited data in the EHR due to patients' variable clinic visits and the lack of standardization of the blood pressure measurement. In addition, the study's inclusion criteria may not have been stringent enough to identify patients with poorly controlled hypertension who may have benefited from inclusion in this intervention. However, despite these limitations we believe the findings from this study add to the literature on engagement in mHealth interventions among patients who receive care at an FQHC.

Conclusions

The findings from this study indicate that the combination of an automated system that identifies at-risk patients using an EHR in addition to tailored education via mHealth can successfully be used to provide self-management education to a high-risk, low-income population. The findings from this study indicate that population health applications can be easily applied with a potentially promising impact in an FQHC.

Acknowledgments

Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award numbers R34, DK102166-01, and P30DK096493. UDP was also supported by R01DK93938 and P30DK096493 prior to joining Gilead Sciences in 2016. CJD was supported by grants funded by the National Institute of Diabetes and Digestive and Kidney Diseases (K23-DK099385 and R01-DK093938). HBB was supported by a Research Career Scientist award from Veterans Affairs (VA) Health Services Research and Development (HSR&D 08-027). Support for AAL was provided by the VA Office of Academic Affiliations (TPH 21-000), and publication support was provided by Durham VA Health Services Research Center of Innovation funding (CIN 13-410). The content is solely the responsibility of the authors and does not necessarily reflect the position or policy of Duke University, the US Department of Veterans Affairs, or the US government.

Conflicts of Interest

HBB reports receiving research funds from Sanofi, Otsuka, Johnson and Johnson, and Improved Patient Outcomes as well as consulting funds from Sanofi. UDP is employed by Gilead Sciences, Inc.

References

1. Losito A, Pittavini L, Zampi I, Zampi E. Characteristics of the relationship of kidney dysfunction with cardiovascular disease in high risk patients with diabetes. Int J Nephrol 2016;2016:7180784 [FREE Full text] [doi: 10.1155/2016/7180784] [Medline: 27885341]
2. Leehey DJ, Zhang JH, Emanuele NV, Whaley-Connell A, Palevsky PM, Reilly RF, VA NEPHRON-D Study Group. BP and renal outcomes in diabetic kidney disease: the Veterans Affairs Nephropathy in Diabetes Trial. Clin J Am Soc Nephrol 2015 Dec 07;10(12):2159-2169 [FREE Full text] [doi: 10.2215/CJN.02850315] [Medline: 26482258]
3. Hill CJ, Cardwell CR, Patterson CC, Maxwell AP, Magee GM, Young RJ, et al. Chronic kidney disease and diabetes in the National Health Service: a cross-sectional survey of the U.K. national diabetes audit. Diabet Med 2014 Apr;31(4):448-454. [doi: 10.1111/dme.12312] [Medline: 24102856]
4. National Diabetes Statistics Report, 2017: estimates of diabetes and its burden in the United States.: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2017. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2018-01-31] [WebCite Cache ID 76YpDvpfg]
5. Kumar S, de Lusignan S, McGovern A, Correa A, Hriskova M, Gatenby P, et al. Ischaemic stroke, haemorrhage, and mortality in older patients with chronic kidney disease newly started on anticoagulation for atrial fibrillation: a population based study from UK primary care. BMJ 2018 Dec 14;360:k342. [doi: 10.1136/bmj.k342] [Medline: 29444881]
6. Chang P, Chien L, Lin Y, Chiu W, Chiou H. Simultaneous control of glycemic, blood pressure, and lipid significantly reduce the risk of renal progression in diabetes patients. Eur J Intern Med 2016 Dec;36:87-92. [doi: 10.1016/j.ejim.2016.09.013] [Medline: 27707566]

7. Duru OK, Middleton T, Tewari MK, Norris K. The landscape of diabetic kidney disease in the United States. Curr Diab Rep 2018 Feb 19;18(3):14 [FREE Full text] [doi: 10.1007/s11892-018-0980-x] [Medline: 29457196]
8. Murphy D, McCulloch CE, Lin F, Banerjee T, Bragg-Gresham JL, Eberhardt MS, Centers for Disease Control and Prevention Chronic Kidney Disease Surveillance Team. Trends in prevalence of chronic kidney disease in the United States. Ann Intern Med 2016 Oct 04;165(7):473-481 [FREE Full text] [doi: 10.7326/M16-0273] [Medline: 27479614]
9. Kidney disease statistics for the United States.: National Institute of Diabetes and Digestive and Kidney Diseases; 2016. URL: https://www.niddk.nih.gov/health-information/health-statistics/kidney-disease [accessed 2018-03-06] [WebCite Cache ID 76YopBRDX]
10. Moreau C, Sando KR, Zambrano DH. Assessing the effect of pharmacist care on diabetes-related outcomes in a rural outpatient clinic: a retrospective case-control study. Ann Pharmacother 2017 Jun;51(6):473-478. [doi: 10.1177/1060028016685731] [Medline: 28064508]
11. Young BA, Katz R, Boulware LE, Kestenbaum B, de Boer IH, Wang W, et al. Risk factors for rapid kidney function decline among African Americans: the Jackson Heart Study (JHS). Am J Kidney Dis 2016 Dec;68(2):229-239 [FREE Full text] [doi: 10.1053/j.ajkd.2016.02.046] [Medline: 27066930]
12. Lo C, Ilic D, Teede H, Cass A, Fulcher G, Gallagher M, et al. The perspectives of patients on health-care for co-morbid diabetes and chronic kidney disease: a qualitative study. PLoS One 2016;11(1):e0146615 [FREE Full text] [doi: 10.1371/journal.pone.0146615] [Medline: 26730708]
13. Polonia J, Azevedo A, Monte M, Silva JA, Bertoquini S. Annual deterioration of renal function in hypertensive patients with and without diabetes. Vasc Health Risk Manag 2017;13:231-237 [FREE Full text] [doi: 10.2147/VHRM.S135253] [Medline: 28721063]
14. Lovre D, Shah S, Sihota A, Fonseca VA. Managing diabetes and cardiovascular risk in chronic kidney disease patients. Endocrinol Metab Clin North Am 2018 Dec;47(1):237-257. [doi: 10.1016/j.ecl.2017.10.006] [Medline: 29407054]
15. Anderson RJ, Bahn GD, Emanuele NV, Marks JB, Duckworth WC, VADT Study Group. Blood pressure and pulse pressure effects on renal outcomes in the Veterans Affairs Diabetes Trial (VADT). Diabetes Care 2014 Oct;37(10):2782-2788 [FREE Full text] [doi: 10.2337/dc14-0284] [Medline: 25048382]
16. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
17. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
18. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. BMJ 2018 May 17;361:k1959. [doi: 10.1136/bmj.k1959]
19. Mobile fact sheet. Washington: Pew Internet and American Life Project; 2018. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-04-10] [WebCite Cache ID 76Yocmarv]
20. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. 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]
21. Kruse RL, Koopman RJ, Wakefield BJ, Wakefield DS, Keplinger LE, Canfield SM, et al. Internet use by primary care patients: where is the digital divide? Fam Med 2012 May;44(5):342-347 [FREE Full text] [Medline: 23027117]
22. López L, Green AR, Tan-McGrory A, King R, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011 Oct;37(10):437-445. [Medline: 22013816]
23. Lee ES, Vedanthan R, Jeemon P, Kamano JH, Kudesia P, Rajan V, et al. Quality improvement for cardiovascular disease care in low- and middle-income countries: a systematic review. PLoS One 2016;11(6):e0157036 [FREE Full text] [doi: 10.1371/journal.pone.0157036] [Medline: 27299563]
24. 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;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
25. Diamantidis CJ, Bosworth HB, Oakes MM, Davenport CA, Pendergast JF, Patel S, et al. Simultaneous Risk Factor Control Using Telehealth to Slow Progression of Diabetic Kidney Disease (STOP-DKD) study: protocol and baseline characteristics of a randomized controlled trial. Contemp Clin Trials 2018 Jun;69:28-39. [doi: 10.1016/j.cct.2018.04.003] [Medline: 29649631]
26. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 9;288(14):1775-1779. [Medline: 12365965]
27. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 2002 Oct 16;288(15):1909-1914. [Medline: 12377092]

28. Bosworth HB, Oddone EZ. A model of psychosocial and cultural antecedents of blood pressure control. J Natl Med Assoc 2002 Apr;94(4):236-248. [Medline: 11991336]
29. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
30. Renders CM, Valk GD, Griffin SJ, Wagner EH, Eijk VJT, Assendelft WJ. Interventions to improve the management of diabetes in primary care, outpatient, and community settings: a systematic review. Diabetes Care 2001 Oct;24(10):1821-1833. [Medline: 11574449]
31. Haynes RB, McDonald H, Garg AX, Montague P. Interventions for helping patients to follow prescriptions for medications. Cochrane Database Syst Rev 2002(2):CD000011. [doi: 10.1002/14651858.CD000011] [Medline: 12076376]
32. Sarkar U, Handley MA, Gupta R, Tang A, Murphy E, Seligman HK, et al. Use of an interactive, telephone-based self-management support program to identify adverse events among ambulatory diabetes patients. J Gen Intern Med 2008 Apr;23(4):459-465 [FREE Full text] [doi: 10.1007/s11606-007-0398-7] [Medline: 18373145]
33. Bosworth H, Oddone E, Weinberger M. Patient Treatment Adherence: Concepts, Interventions, and Measurement. Mahwah: Lawrence Erlbaum Associates; 2006.
34. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975 Jan;13(1):10-24. [Medline: 1089182]
35. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Advances Behav Res Ther 1978 Jan;1(4):139-161. [doi: 10.1016/0146-6402(78)90002-4]
36. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
37. Khalili M, Wong RJ. Underserved does not mean undeserved: unfurling the HCV care in the safety net. Dig Dis Sci 2018 Dec;63(12):3250-3252. [doi: 10.1007/s10620-018-5316-9] [Medline: 30311153]
38. Agresti A. Categorical Data Analysis. Hoboken: Wiley-Interscience; 2013.
39. Vaghefi I, Hughes JB, Law S, Lortie M, Leaver C, Lapointe L. Understanding the impact of electronic medical record use on practice-based population health management: a mixed-method study. JMIR Med Inform 2016 Apr 04;4(2):e10 [FREE Full text] [doi: 10.2196/medinform.4577] [Medline: 27044411]
40. Schmittdiel JA, Gopalan A, Lin MW, Banerjee S, Chau CV, Adams AS. Population health management for diabetes: health care system-level approaches for improving quality and addressing disparities. Curr Diab Rep 2017 Dec;17(5):31 [FREE Full text] [doi: 10.1007/s11892-017-0858-3] [Medline: 28364355]
41. Ishani A, Greer N, Taylor BC, Kubes L, Cole P, Atwood M, et al. Effect of nurse case management compared with usual care on controlling cardiovascular risk factors in patients with diabetes: a randomized controlled trial. Diabetes Care 2011 Aug;34(8):1689-1694 [FREE Full text] [doi: 10.2337/dc10-2121] [Medline: 21636796]
42. Turchin A, Goldberg SI, Shubina M, Einbinder JS, Conlin PR. Encounter frequency and blood pressure in hypertensive patients with diabetes mellitus. Hypertension 2010 Jul;56(1):68-74 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.109.148791] [Medline: 20497991]
43. Helou N, Dwyer A, Shaha M, Zanchi A. Multidisciplinary management of diabetic kidney disease: a systematic review and meta-analysis. JBI Database System Rev Implement Rep 2016 Dec;14(7):169-207. [doi: 10.11124/JBISRIR-2016-003011] [Medline: 27532796]
44. Einstadter D, Bolen SD, Misak JE, Bar-Shain DS, Cebul RD. Association of repeated measurements with blood pressure control in primary care. JAMA Intern Med 2018 Jun 01;178(6):858-860 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0315] [Medline: 29710186]
45. Choi S, Shin J, Choi SY, Sung KC, Ihm SH, Kim K, et al. Impact of visit-to-visit variability in systolic blood pressure on cardiovascular outcomes in Korean National Health Insurance Service-National Sample Cohort. Am J Hypertens 2017 Jun 01;30(6):577-586. [doi: 10.1093/ajh/hpw157] [Medline: 28052880]
46. Wu C, Shlipak MG, Stawski RS, Peralta CA, Psaty BM, Harris TB, Health ABC Study. Visit-to-visit blood pressure variability and mortality and cardiovascular outcomes among older adults: the health, aging, and body composition study. Am J Hypertens 2017 Feb;30(2):151-158 [FREE Full text] [doi: 10.1093/ajh/hpw106] [Medline: 27600581]

Abbreviations

BP: blood pressure
DBP: diastolic blood pressure
DKD: diabetic kidney disease
EHR: electronic health record
FQHC: federally qualified health center
ICD: International Classification of Diseases
mHealth: mobile health
SBP: systolic blood pressure

STOP-DKD APP: Simultaneous Risk Factor Control Using Telehealth to Slow Progression of Diabetic Kidney Disease Automated Population Program
VA: Veterans Affairs

Edited by G Eysenbach; submitted 18.10.18; peer-reviewed by CS Hung, D Robinson; comments to author 12.12.18; revised version received 26.01.19; accepted 27.01.19; published 09.04.19.

Please cite as:
Lewinski AA, Patel UD, Diamantidis CJ, Oakes M, Baloch K, Crowley MJ, Wilson J, Pendergast J, Biola H, Boulware LE, Bosworth HB
Addressing Diabetes and Poorly Controlled Hypertension: Pragmatic mHealth Self-Management Intervention
J Med Internet Res 2019;21(4):e12541
URL: https://www.jmir.org/2019/4/e12541/
doi:10.2196/12541
PMID:30964439

Original Paper

Effectiveness of Internet-Delivered Computerized Cognitive Behavioral Therapy for Patients With Insomnia Who Remain Symptomatic Following Pharmacotherapy: Randomized Controlled Exploratory Trial

Daisuke Sato[1,2], OTR, PhD; Naoki Yoshinaga[3], RN, PhD; Eiichi Nagai[4], PharmD; Kazue Nagai[5], PhD; Eiji Shimizu[1,6,7], MD, PhD

[1]Department of Cognitive Behavioral Physiology, Graduate School of Medicine, Chiba University, Chiba, Japan
[2]Department of Rehabilitation Sciences, Chiba Prefectural University of Health Sciences, Chiba, Japan
[3]Organization for Promotion of Tenure Track, University of Miyazaki, Miyazaki, Japan
[4]Clinical Research Center, Chiba University Hospital, Chiba, Japan
[5]Research and Education Center of Health Sciences, Gunma University Graduate School of Health Sciences, Maebashi, Japan
[6]Research Center for Child Mental Development, Chiba University, Chiba, Japan
[7]Cognitive Behavioral Therapy Center, Chiba University Hospital, Chiba, Japan

Corresponding Author:
Daisuke Sato, OTR, PhD
Department of Cognitive Behavioral Physiology
Graduate School of Medicine
Chiba University
1-8-1
Inohana, Chuo-ku
Chiba, 260-8670
Japan
Phone: 81 43 226 2027
Fax: 81 43 226 2028
Email: daisuke-sato@umin.ac.jp

Abstract

Background: In reality, pharmacotherapy still remains the most common treatment for insomnia.

Objective: This study aimed to examine the effectiveness of our internet-delivered computerized cognitive behavioral therapy (ICBT) program as an adjunct to usual care (UC) compared with UC alone in patients with insomnia who remain symptomatic following hypnotics.

Methods: We recruited 23 patients with insomnia who remained symptomatic following pharmacologic treatment including benzodiazepines, and we conducted an exploratory randomized controlled trial. The primary outcome was the Pittsburgh Sleep Quality Index (PSQI) at week 6 of the treatment. Secondary outcomes were sleep onset latency, total sleep time, sleep efficiency, number of awakenings, refreshment and soundness of sleep, anxiety by Hospital Anxiety and Depression Scale, depression measured by the Center for Epidemiologic Studies Depression Scale, and quality of life (QOL) measured by the EuroQol-5D. All parameters were measured at weeks 0 (baseline), 6 (postintervention), and 12 (follow-up).

Results: The adjusted mean reduction (−6.11) in PSQI at week 6 from baseline in the ICBT plus UC group was significantly (P<.001) larger than the adjusted mean reduction (0.40) in the UC alone group. Significant differences were also found in favor of ICBT plus UC for PSQI, sleep onset latency, sleep efficiency, number of awakenings, and depression at all assessment points. Refreshment, soundness of sleep, anxiety, and QOL improved by week 6 in ICBT plus UC compared with UC alone. There were no reports of adverse events in either group during the study.

Conclusions: These results indicated that our 6-week ICBT program is an effective treatment adjunct to UC for improving insomnia and related symptoms even after unsuccessful pharmacotherapy.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry: UMIN000021509; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000023545 (Archived by WebCite at http://www.webcitation.org/75tCmwnYt).

(J Med Internet Res 2019;21(4):e12686) doi:10.2196/12686

KEYWORDS

insomnia; cognitive behavioral therapy; randomized controlled trial; internet; benzodiazepines; residual symptoms

Introduction

Background

Insomnia, which affects 10% to 12% of the total population, is characterized by the inability to fall asleep or awakening too early in the morning or during the night, resulting in nonrestorative sleep and decreased daytime functioning [1-4]. The spontaneous improvement of insomnia is low [5]. The treatment options for insomnia are psychotherapy and pharmacotherapy, and the American College of Physicians (ACP) recommends that all adult patients receive cognitive behavioral therapy (CBT) for insomnia as the initial treatment for chronic insomnia disorder. The ACP also recommends that clinicians use a shared decision-making approach, including a discussion with the patient, of the benefits, harms, and costs of the short-term use of medications before deciding whether to add pharmacological therapy in adults with chronic insomnia disorder in whom CBT for insomnia alone was unsuccessful [6]. In 2005, Vallières et al reported that pharmacotherapy before the initiation of CBT appears to be less effective than the combined treatment of pharmacotherapy plus CBT, followed by CBT alone [7]. Their study also revealed that the early introduction of CBT contributes to a maximization of the effect of pharmacotherapy.

Pharmacotherapy remains the most commonly used treatment option for insomnia worldwide [8]. In Japan, especially, CBT for insomnia is not covered by public health insurance at this time (2018), and thus the most common initial treatment for insomnia is a primary care physician's prescription of an insomnia drug such as a benzodiazepine, nonbenzodiazepine, orexin receptor antagonist, or melatonin receptor agonist and antidepressants [9]. Although pharmacotherapy is associated with a high incidence of adverse effects including daytime sleepiness, recurrent insomnia, and drug dependence [10], physicians in Japan often prescribe excess doses of benzodiazepine [11]. A next step treatment in patients with insomnia who remain symptomatic following pharmacotherapy is strongly needed. Clinical practice guidelines suggest CBT, rather than pharmacotherapy, as the initial therapy for patients with insomnia [6,12,13]. Okajima et al showed that face-to-face CBT with a behavioral analysis is more effective than pharmacotherapy for Japanese chronic insomnia patients who are resistant to pharmacological treatment [14]. As Web-based programs are now more accessible and low cost and can be conveniently completed at one's own time and place [15], we have developed an internet-delivered computerized cognitive behavioral therapy (ICBT) program for insomnia and we published a randomized controlled trial (RCT) design [16].

Objectives

We conducted the RCT to examine the effectiveness of ICBT as an adjunct to usual care (UC) compared with UC alone, specifically targeting insomnia patients who remain symptomatic after pharmacotherapy. We hypothesized that among insomnia patients who remain symptomatic after pharmacotherapy, the augmentation with ICBT would be superior to UC alone in improving overall sleep quality, reducing anxiety and depression, and improving the patients' quality of life (QOL).

Methods

Study Design and Participants

Our study protocol has been published [16] and is therefore only summarized here. This was a randomized controlled single-center trial conducted at the academic outpatient clinic of the Cognitive Behavioral Therapy Center of Chiba University Hospital between March 2016 and January 2018 as the recruitment period and between March 2016 and April 2018 as the trial period. Participants were recruited through posters and leaflets placed at medical institutions in the Chiba prefecture and through Web-based and newspaper advertisements. The inclusion criteria for this study were as follows: the participant regularly went to bed between 8 pm and 2 am; aged 18 to 65 years; having a primary diagnosis of insomnia according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [1]; and the aforementioned insomnia remaining symptomatic. *Remaining symptomatic* was defined herein as having insomnia that is at least moderate in severity, based on a Pittsburgh Sleep Quality Index (PSQI) score of greater than 5.5 [15,17,18] after the use of hypnotics including nonbenzodiazepines, benzodiazepines, melatonin receptor agonists, orexin receptor antagonists, and antidepressants.

Each participant's treatment history was confirmed by their prescribing clinician and by chart review. All patients were evaluated by 2 researchers (a psychiatrist, ES, and a therapist, DS) who also verified the patient diagnosis and eligibility. They discussed the validity of the patient's initial diagnosis and eligibility. Patients were reevaluated to cover important missing information based on suggestions derived from the discussion, and the final diagnosis and eligibility were confirmed by the 2 researchers.

The exclusion criteria included severe symptoms of anxiety or depression. Anxiety was assessed using the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS) which contains 7 items. Depression was assessed using the total score of the Center for Epidemiological Studies Depression (CES-D) scale. Patients with a HADS score of greater than or equal to 10 or a CES-D score of greater than or equal to 30 were excluded. Patients with psychosis, organic mental disorder, or current high risk of suicide, substance abuse, or dependence within the 12 months before enrollment, antisocial personality disorder, or unstable medical condition were also excluded.

Randomization

At the end of the baseline assessment, eligible participants were randomly assigned to either the UC arm or ICBT plus UC arm at a ratio of 1:1, with assignments made using the minimization method, ensuring a balance in baseline PSQI scores (PSQI<12) and gender. Each participant was then assigned to one of the 2 treatment regimes. Participants were blinded to the group to which they were assigned before consenting to participate in the study.

Procedures

Primary physicians referred patients to the trial but continued to provide pharmacotherapy as UC to the patients in both groups, as described [19,20]. As part of the UC, both the UC only and ICBT plus UC groups received email magazines with general information about insomnia and hypnotics (in PDF format) by our research team 4 times over a 6-week period.

The ICBT program for insomnia was developed by one of the authors (ES) and is named the *Insomnia Improvement Internet Program*. The program is called IIIP (pronounced *three P*) for short, as III indicates the Roman numeral and *three P* sounds like *sleepy*.

The ICBT treatment consists of 5 weekly lessons and includes various elements that are commonly incorporated in face-to-face CBT for insomnia as follows: (1) keeping a sleep diary and understanding sleep hygiene; (2) changing sleep-related behaviors, including stimulus control; (3) restructuring distorted beliefs about sleep and sleep-related worries; (4) sleep restriction to increase sleep efficiency(SE); and (5) relaxation training, including breathing exercises and progressive muscle relaxation. Participants completed the 5 lessons over a 6-week period to provide sufficient time to become accustomed to the CBT. One of the authors (DS), a cognitive behavioral therapist, sent weekly emails to the participants to ask them about their homework and progress. The intervention was implemented as a cognitive behavioral therapist supported ICBT. Participants in the control group were offered the ICBT after the trial, if the UC did not make them sleep better.

Outcomes

Primary Outcome

The primary outcome was the change in the PSQI score at week 6 from baseline (week 0). The PSQI is a self-rated questionnaire consisting of 19 questions across 7 subscales (sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbance, use of hypnotics, and daytime dysfunction). Each subscale is scored on a scale of 0 to 3. Subscale scores are summed to a total score ranging from 0 (good quality of sleep) to 21 (very poor quality of sleep). The PSQI was verified as a reliable and valid measure of subjective sleep quality in clinical practice and experimental research [17,18].

Secondary Outcomes

The secondary outcomes included the change in the PSQI score at week 3 and at week 12 from baseline (week 0) and the sleep onset latency (SOL), total sleep time (TST), SE, and number of awakenings (NA) extracted from the PSQI, as well as the current feeling of refreshment, perceived soundness of sleep (assessed by a visual analog scale), the anxiety subscale of HADS measuring anxiety, the CES-D score measuring depression, and EuroQol-5D (EQ-5D) score measuring the participant’s QOL [21].

The total score on the 7 HADS anxiety subscale items ranges from 0 (no symptoms of anxiety) to 21 (severe symptoms of anxiety). The total score on the 20 CES-D items ranges from 0 (no symptoms of depression) to 60 (severe symptoms of depression). We have described the 3-level version of the EQ-5D [21]. The EQ-5D [21] contains 5 items that assess QOL on a 3-point Likert scale ranging from 1 (not severe) to 3 (severe). The Japanese version of the EQ-5D was developed by Tsuchiya et al [22]. The EQ-5D is the most commonly used scale worldwide for calculating quality-adjusted life years (QALYs). QALYs are often used as the health outcome in cost-utility analyses and are typically estimated via an area under the curve analysis which involves summing the areas of the distribution shapes to calculate utility scores over the study period [23,24]. Our participants completed the HADS, CES-D, and EQ-5D questionnaires at home and sent them to us by email.

The therapist asked the participants about adverse event experiences at each assessment. All measures were assessed at weeks 0 (baseline), 3 (midintervention), 6 (postintervention), and 12 (follow-up).

Statistical Analyses

The statistical analyses and reporting of this trial were conducted in accordance with the Consolidated Standards of Reporting Trials (CONSORT) guidelines. For baseline variables, summary statistics were constructed, using frequencies and proportions for categorical data and the mean and SD for continuous variables. Baseline variables were compared using Fisher exact test for categorical outcomes and the unpaired t test for continuous variables. For the primary analysis comparing treatment effects, the least-squares means, and their 95% CIs were estimated by an analysis of covariance (ANCOVA) with the change in total PSQI scores at week 6. This ANCOVA model took into account the variation caused by treatment effects, and the participants’ gender and baseline PSQI score were entered as covariates. Analyses of secondary outcomes were performed in the same manner as the primary analysis. All P values were 2-sided. P values less than .05 were considered significant. All statistical analyses were performed using SAS version 9.4 software (SAS Institute).

As described by our published design study [16], the sample size was based on a previous study by van Straten et al [25], which indicated that the estimated group difference in changes of PSQI scores from baseline was approximately 2.86 (ICBT group=3.00; control group=0.04). Assuming a group difference of 2.86 points (SD 2.5), 13 subjects per group will provide 80% power to detect a difference in PSQI scores between the UC arm and ICBT plus UC arm, using a 2-sided, 2-sample *t* test at a 5% significance level. Thus, allowing for a 10% dropout rate, 15 participants are required per group, for a total of 30 participants in the study.

Ethical Approval

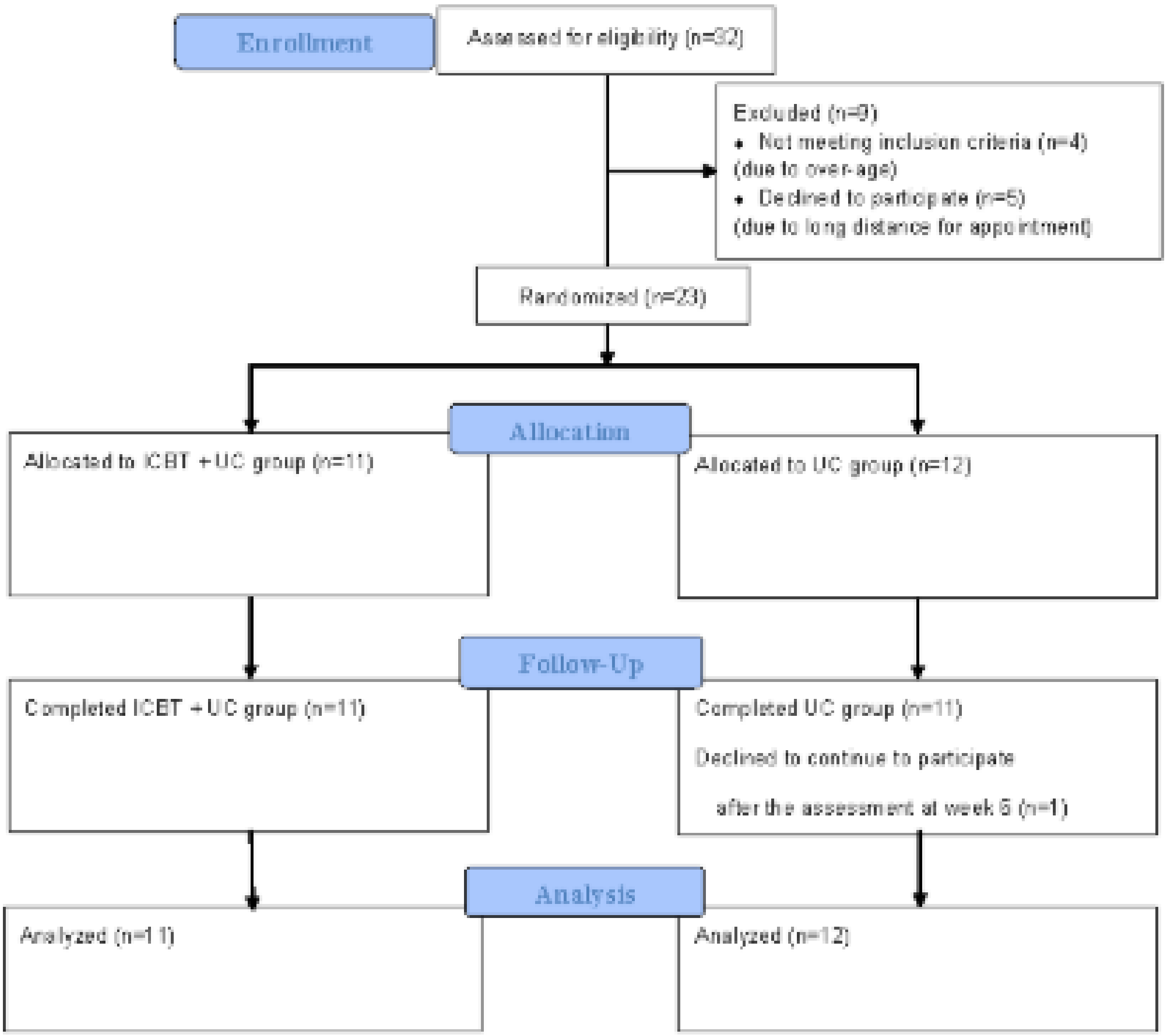
Written informed consent was obtained from all patients after the procedures had been fully explained. Ethical approval was obtained from the Institutional Review Board of Chiba University Hospital (no. G27040), and the trial was registered as UMIN000021509.

Results

Recruitment

Figure 1 shows the patient recruitment flow diagram, based on the CONSORT guidelines. A total of 32 patients applied to participate through our website. Of the 9 patients who were excluded, 4 did not meet one of the inclusion criteria because of their age (over the limit) and 5 declined to participate because of the long distance to our hospital. A final total of 23 patients attended the face-to-face baseline assessment, and all 23 were enrolled in the study. We randomly assigned the 23 patients to the ICBT plus UC and UC groups. Furthermore, 1 patient in the UC group declined to continue to participate and dropped out from the study after the assessment at week 6. Though the originally planned recruitment rate would be 2 participants per month, the real average recruitment rate was 1 participant per month through posters and leaflets placed at medical institutions in the Chiba prefecture and through Web-based and newspaper advertisements within the planned recruitment period between March 2016 and January 2018. We had to stop recruitment after we entered 23 patients after 23 months of trial commencement on the closing date (see Figure 1).

Figure 1. The CONSORT flow diagram for the trial. ICBT: internet-delivered computerized cognitive behavioral therapy; UC: usual care.



Demographics and Clinical Characteristics

Table 1 summarizes the baseline demographic and clinical characteristics of the participants. There were no significant between-group differences in any of the characteristics, including the number of patients with a baseline PSQI score less than 12 (PSQI<12) and the total PSQI score. The ICBT plus UC group received pharmacologic treatment including zolpidem (n=3), brotizolam (n=2), eszopiclone (n=1), etizolam (n=1), estazolam (n=1), flunitrazepam (n=1), alprazolam (n=1), and the combined use of zolpidem and brotizolam (n=1). The UC group received pharmacologic treatment including zolpidem (n=3), brotizolam (n=2), zopiclone (n=1), etizolam (n=1), the combined uses of zolpidem and etizolam (n=1); lormetazepam and flunitrazepam (n=1); quazepam and lorazepam (n=1); brotizolam, nitrazepam, and clonazepam (n=1); and zolpidem, lormetazepam, and trazodone hydrochloride (n=1). Thus, all 23 of the participants were taking one or more benzodiazepines. There was no change of use of sleep medication during intervention and follow-up period among the ICBT plus UC and UC group patients. We are conducting further study to estimate a difference in reduced use of sleep medication after the end of the follow-up period among the ICBT plus UC and UC group patients.

Table 1. Baseline characteristics (N=23).

Variable	ICBT[a] + UC[b] (n=11)	UC (n=12)
Female, n (%)	9 (81.8)	9 (75.0)
Age in years, mean (SD)	49.4 (13.8)	50.5 (8.8)
Material status (married of living as married), n (%)	9 (81.8)	10 (83.3)
Length of education in years, mean (SD)	14.1 (2.5)	14.5 (1.9)
Employment status (in paid employment, full or part-time), n (%)	9 (81.7)	9 (75.0)
Alcohol drinking (habitual or opportunity drinking), n (%)	3 (27.3)	7 (58.3)
Smoking, n (%)	0 (0)	0 (0)
Duration of insomnia in years, mean (SD)	6.3 (5.1)	6.0 (7.7)
Number of patients with PSQI[c] less than 12, n (%)	3 (27.3)	3 (25.0)

[a] ICBT: internet-delivered computerized cognitive behavioral therapy.

[b] UC: usual care.

[c] PSQI: Pittsburgh Sleep Quality Index.

Primary Outcome

The raw data (ie, mean and SDs) of the participants' PSQI scores at the 4 assessment points are shown in Figure 2 and Table 2. At week 6, the adjusted mean reductions in PSQI from baseline were −6.11 (95% CI −7.45 to −4.78) and 0.40 (95% CI −0.83 to 1.63) for the ICBT plus UC and UC groups, respectively. The group difference was significant at −6.51 (95% CI −8.15 to 4.87, *P*<.001; Table 3). The combination therapy, that is, ICBT plus UC, was therefore superior to UC alone.

Secondary Outcomes

At week 3, the adjusted mean reductions in PSQI from baseline were −2.66 (95% CI −3.63 to −1.69) and 0.45 (95% CI −0.45 to 1.34) for the ICBT plus UC and UC groups, respectively. The group difference was significant at −3.10 (95% CI −4.29 to 1.92, *P*<.001; Table 3). At week 12, the adjusted mean reductions in PSQI from baseline were −6.40 (95% CI −8.05 to −4.75) and 0.44 (95% CI −1.11 to 1.99) for the ICBT plus UC and UC groups, respectively. The group difference was significant at −6.84 (95% CI −8.90 to 4.77, *P*<.001; Table 3). The combination therapy of ICBT plus UC was thus superior to UC alone at all 3 of the assessment time points.

The raw data (mean and SDs) and the adjusted mean changes for the secondary outcome measures are presented in Table 2 and Table 3, respectively. At week 3, compared with the UC group, significant improvements were observed in the ICBT plus UC group in the SOL, SE, NA scores, and CES-D (all *P*<.05). At week 6, compared with the UC group, significant improvements were observed in the ICBT plus UC group in the SOL, SE, NA, current feeling of refreshment, perceived soundness of sleep, the anxiety subscale of HADS, the CES-D, and the EQ-5D (all *P*<.05). At week 12, compared with the UC group, significant improvements were observed in the ICBT plus UC group in SOL, SE, NA, current feeling of refreshment, perceived soundness of sleep, the anxiety subscale of HADS, the CES-D, and the EQ-5D (all *P*<.05). There were no significant differences in the TST between the 2 groups at the 3 time points.

This result shows that, compared with the insomnia patients who received only UC, those who received ICBT plus UC reported significant improvements in their current feeling of refreshment, perceived soundness of sleep, measures of anxiety and depression, and functioning or QOL improved by week 6. The administration of ICBT improved the SOL, SE, NA, and depression at an early stage, but not the TST at the final stage.

Figure 2. Mean and SDs (raw data) for the primary outcome, the Pittsburgh Sleep Quality Index (PSQI) score improvement. ICBT: internet-delivered computerized cognitive behavioral therapy; UC: usual care.

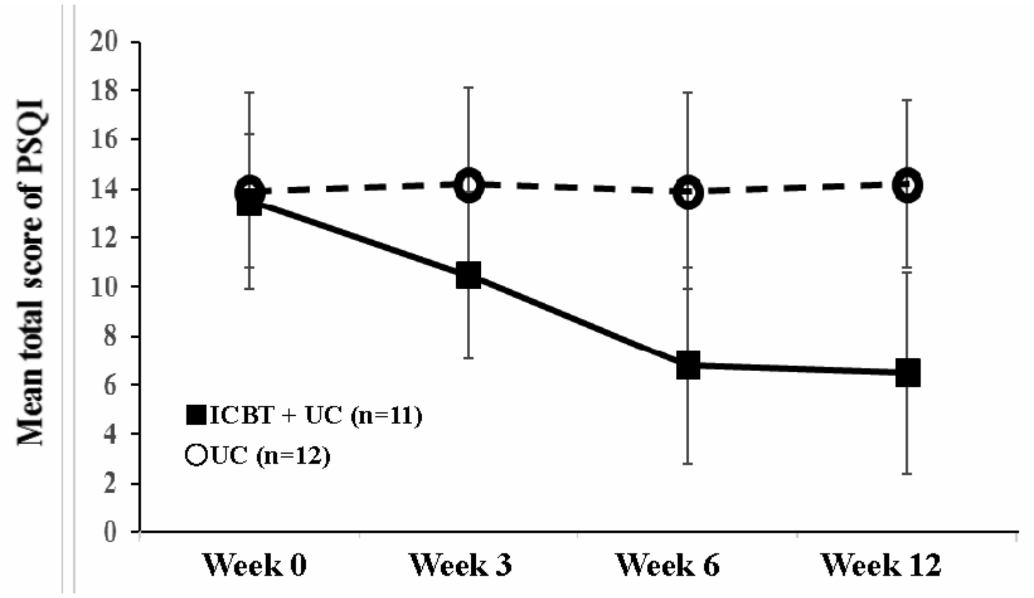


Table 2. Raw data of the primary and secondary outcomes (N=23).

Variable	ICBT[a] + UC[b] (n=11)	UC (n=12)
Primary outcome: sleep characteristics		
PSQI[c], mean (SD)	13.5 (2.7)	13.9 (4.0)
Secondary outcomes: sleep		
SOL[d], minute, mean (SD)	54.5 (43.4)	41.7 (20.7)
TST[e], hour, mean (SD)	6.3 (2.1)	5.9 (1.2)
Sleep efficiency, %, mean (SD)	73.2 (8.4)	73.6 (20.6)
NA[f], mean (SD)	3.5 (2.9)	2.9 (1.7)
Refreshment, mean (SD)	3.8 (2.6)	4.3 (2.5)
Soundness of sleep, mean (SD)	4.0 (2.7)	3.2 (2.2)
Secondary outcomes: health		
Anxiety, HADS[g], mean (SD)	4.8 (3.2)	6.7 (2.5)
Depression, CES-D[h], mean (SD)	18.5 (4.6)	21.4 (5.4)
QOL[i], EQ-5D[j], mean (SD)	0.911 (0.125)	0.703 (0.174)

[a] ICBT: internet-delivered computerized cognitive behavioral therapy.

[b] UC: usual care.

[c] PSQI: Pittsburgh Sleep Quality Index.

[d] SOL: sleep onset latency.

[e] TST: total sleep time.

[f] NA: number of awakenings.

[g] HADS: Hospital Anxiety and Depression Scale.

[h] CES-D: Center for Epidemiologic Studies Depression Scale.

[i] QOL: quality of life.

[j] EQ-5D: EuroQol-5D.

Table 3. Adjusted mean changes in the primary and secondary outcomes.

Changes from baseline	ICBT[a]+UC[b] (n=11)		UC (n=12)		Intergroup difference		*P* value
	Least squares mean	95% CI	Least squares mean	95% CI	Difference	95% CI	
Primary outcome, PSQI[c], week 6	−6.11	−7.45 to −4.78	0.40	−0.83 to 1.63	−6.51	−8.15 to −4.87	<.001
Secondary outcomes							
PSQI							
Week 3	−2.66	−3.63 to −1.69	0.45	−0.45 to 1.34	−3.10	−4.29 to −1.92	<.001
Week 12	−6.40	−8.05 to −4.75	0.44	−1.11 to 1.99	−6.84	−8.90 to 4.77	<.001
SOL[d], minute							
Week 3	−10.82	−19.58 to −2.06	3.25	−5.00 to 11.49	−14.07	−24.97 to −3.16	.01
Week 6	−27.01	−34.83 to −19.19	4.64	−2.73 to 12	−31.65	−41.39 to −21.91	<.001
Week 12	−29.32	−38.90 to −19.74	6.47	−2.73 to 15.66	−35.79	−47.88 to −23.70	<.001
TST[e], hour							
Week 3	0.49	0.13 to 0.86	0.07	−0.27 to 0.41	0.43	−0.03 to −0.88	.06
Week 6	0.58	0.03 to 1.13	−0.08	−0.59 to 0.43	0.66	−0.02 to 1.34	.06
Week 12	0.49	−0.08 to 1.05	−0.05	−0.59 to 0.48	0.54	−0.17 to 1.25	.13
Sleep efficiency, %							
Week 3	4.65	0.59 to 8.72	−3.72	−7.45 to 0.01	8.37	3.44 to 13.31	.002
Week 6	13.24	8.67 to 17.81	−4.68	−8.87 to −0.48	17.92	12.37 to 23.47	<.001
Week 12	13.32	8.68 to 17.97	−7.00	−11.37 to −2.63	20.32	14.55 to 26.1	<.001
NA[f]							
Week 3	−0.53	−1.17 to 0.12	0.61	0.01 to 1.21	−1.14	−1.93 to −0.34	.008
Week 6	−2.04	−2.83 to −1.26	0.17	−0.55 to 0.89	−2.22	−3.18 to −1.25	<.001
Week 12	−1.95	−2.81 to −1.08	0.05	−0.77 to 0.87	−2.00	−3.08 to −0.91	<.001
Refreshment							
Week 3	0.53	−0.30 to 1.35	−0.20	−0.95 to 0.56	0.72	−0.28 to 1.73	.15
Week 6	2.09	0.67 to 3.51	−0.49	−1.79 to 0.81	2.58	0.84 to 4.31	.006
Week 12	2.06	0.48 to 3.64	−1.03	−2.52 to 0.46	3.09	1.11 to 5.07	.004
Soundness of sleep							
Week 3	0.12	−0.82 to 1.06	−0.61	−1.49 to 0.27	0.73	−0.44 to 1.9	.20
Week 6	2.76	1.25 to 4.28	0.30	−1.11 to 1.72	2.46	0.58 to 4.34	.01
Week 12	2.70	1.21 to 4.18	−0.36	−1.77 to 1.06	3.05	1.15 to 4.96	.004
Anxiety, HADS[g]							
Week 3	−0.08	−0.74 to 0.58	0.68	0.11 to 1.26	−0.77	−1.58 to 0.04	.06
Week 6	−0.83	−1.62 to −0.04	1.17	0.48 to 1.85	−1.99	−2.96 to −1.02	<.001
Week 12	−0.95	−1.98 to 0.08	1.45	0.54 to 2.37	−2.40	−3.70 to −1.11	<.001
Depression, CES-D[h]							
Week 3	−2.46	−4.34 to −0.59	0.74	−1.00 to 2.47	−3.20	−5.56 to −0.84	.01
Week 6	−5.18	−7.77 to −2.60	1.14	−1.25 to 3.53	−6.32	−9.57 to −3.07	<.001
Week 12	−5.94	−8.92 to −2.96	2.02	−0.82 to 4.85	−7.96	−11.77 to −4.14	<.001
QOL[i], EQ-5D[j]							

Changes from baseline	ICBT[a]+UC[b] (n=11)		UC (n=12)		Intergroup difference		*P* value
	Least squares mean	95% CI	Least squares mean	95% CI	Difference	95% CI	
Week 3	−0.0065	−0.0516 to 0.0386	−0.0145	−0.0546 to 0.0256	0.0079	−0.0529 to 0.0688	.78
Week 6	0.0487	0.0001 to 0.0972	−0.0793	−0.1225 to −0.0361	0.128	0.0625 to 0.1935	<.001
Week 12	0.059	0.003 to 0.115	−0.0943	−0.1457 to −0.0429	0.1533	0.0759 to 0.2307	<.001

[a] ICBT: internet-delivered computerized cognitive behavioral therapy.

[b] UC: usual care.

[c] PSQI: Pittsburgh Sleep Quality Index.

[d] SOL: sleep onset latency.

[e] TST: total sleep time.

[f] NA: number of awakenings.

[g] HADS: Hospital Anxiety and Depression Scale.

[h] CES-D: Center for Epidemiologic Studies Depression Scale.

[i] QOL: quality of life.

[j] EQ-5D: EuroQol-5D.

For the screening of insomnia, 5.5 is considered the optimal cut-off score of the PSQI [26,27]. We therefore set the threshold at a PSQI score of 5.5 to determine the remission of insomnia. At week 6, 36% (4/11) of the patients in the ICBT plus UC group showed remission of insomnia with a PSQI score less than 5.5, whereas no UC group patient showed remission. At week 12, 45% (5/11) of the patients in the ICBT plus UC group and no patient in the UC group showed remission with a PSQI score of less than 5.5. The remission rates at week 6 and 12 were significantly higher in the ICBT plus UC group compared with the UC group by Fisher exact test (*P*<.05). There were no participants who could not complete the 5 lessons over a 6-week period in the intervention group and did not get back to a cognitive behavioral therapist's weekly emails about their homework and progress. There were no reports of any adverse events in either group during the study.

Discussion

Principal Findings

This is the first RCT to examine the effectiveness of ICBT as a next step treatment for patients with insomnia who remain symptomatic despite drug treatment. Our findings demonstrate that the ICBT was effective as an adjunct to UC in reducing the severity of insomnia at week 6 immediately after the intervention. Moreover, the patients who received ICBT showed significant improvements at week 3 as a midpoint of the intervention and at week 12 of the follow-up period.

Comparison With Prior Work

A meta-analysis of RCTs including 14 records of 15 studies (1013 experimental group participants and 591 waiting list participants) showed that internet-based CBT for adults with insomnia is an effective treatment [28]. Except for being insomnia patients who remain symptomatic following treatment with a hypnotic in this study, the baseline clinical characteristics (age and sex) of our recruited patients in Japan are similar to those in Western countries. In their study of patients recruited from the general population, van Straten et al [25] reported that their guided ICBT for insomnia changed the mean PSQI score from 12.4 (SD 2.1) at pretreatment to 8.9 (SD 2.6) at week 6 post-treatment compared with a wait-list control group's score reduction from 11.7 (SD 2.2) to 11.6 (SD 2.5). In an investigation of patients with comorbid psychiatric diagnoses who were taking one or more psychotropic medications, Feuerstein et al [29] showed that their computer-based delivery of CBT for insomnia significantly improved the patients' PSQI scores compared with an active control group (sleep diary group). Our present findings regarding PSQI improvement seem comparable to these 2 studies even though the patient populations differ.

A meta-analysis of RCTs including 87 RCTs comparing 118 treatments (3724 patients) to nontreated controls (2579 patients) showed that face-to-face and ICBT for adults with insomnia are effective treatments [30]. In addition, that meta-analysis described between-group effect sizes of outcomes concerning sleep as follows: insomnia severity index (Hedges g=0.98), SE (g=0.71), PSQI (g=0.65), wake after sleep onset (g=0.63) and SOL (g=0.57), NA (g=0.29), and sleep quality (g=0.40). The meta-analysis authors also mentioned that the smallest effect was on TST (g = 0.16) [30]. The various effect sizes on different sleep outcomes seem to be consistent with our finding that our ICBT program showed a remarkably large Hedges g value for PSQI (g=−3.36), SE (g=2.36), SOL (g=−1.80), and NA (g=−1.39), but not for TST (g=0.42); we calculated the between-group effect sizes at week 6 from baseline.

Moreover, Lancee et al [31] reported superior performance of face-to-face treatment relative to online treatment in their RCT comparing 3 conditions: guided online, face-to-face, and wait-list. In Japan, Yamadera et al [32] reported that face-to-face individual CBT for insomnia resulted in a PSQI improvement from 12.7 (SD 0.7) to 8.9 (SD 0.6) compared with the improvement because of group CBT from 12.2 (SD 0.5) to 10.1

(SD 0.7). Okajima et al [14] reported that face-to-face individual CBT for pharmacological treatment-resistant chronic insomnia resulted in PSQI improvement from 13.59 (SD 3.25) to 8.10 (SD 2.95) compared with the improvement because of UC from 12.45 (SD 2.52) to 11.17 (SD 3.23). Our observation of PSQI improvement from 13.5 (SD 2.7) to 6.8 (SD 4.0) versus the improvement with UC from 13.9 (SD 4.0) to 13.9 (SD 4.0) seems to be comparable to the above 2 Japanese studies of face-to-face treatment, even though we used a guided online program. Further research is necessary to compare online CBT with face-to-face CBT, including cost-effectiveness and patients' preferences.

According to the algorithm in a new clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults issued by the American Academy of Sleep Medicine [33], if pharmacologic treatment (short-term intermediate-acting benzodiazepine receptor agonists or ramelteon) does not improve the symptoms of an individual with chronic insomnia, clinicians should consider switching to another modality (ie, CBT) or combined treatment with CBT. The results of our present analyses suggest that the simple continuation of pharmacologic treatment was largely ineffective for our population of patients with insomnia and that clinicians should consider providing ICBT or referring patients to a CBT therapist if pharmacologic treatment is not sufficiently effective.

A meta-analysis of adherence to ICBT showed that the percentage of noncompleters of total ICBT intervention was 34.9% [34]. All participants in the ICBT plus UC group accomplished the total ICBT program in our study, one strength of this study is the low rate of dropout.

Limitations

This study has the following 5 limitations. First, we were unable to elucidate specific effects of the ICBT program because a psychological placebo group was not used to control for nonspecific factors. Second, the sample size was relatively small (n=23). Third, the lack of 1-year follow-up data limits the generalizability of our conclusions. Larger- and longer-scale studies are necessary. Fourth, sleep estimates were based on subjective sleep diaries and PSQI scores, rather than on objective measures such as polysomnography. The use of both subjective and objective measures has been recommended [35,36]. Finally, Outcome assessors were not blinded; however, blinded outcome assessment is recommended in open label trials to reduce bias.

Conclusions

In conclusion, our results suggested that a 6-week ICBT program is an effective treatment for patients with insomnia who remain symptomatic following pharmacologic treatment.

Acknowledgments

This study was supported by part of the Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (JSPS) (grant KAKENHI 17H04091 to Eiji Shimizu).

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v21i4e12686_app1.pdf]

References

1. American Psychiatric Association. 5th ed. In: Diagnostic and statistical manual of mental disorders: DSM-5. Washington: American Psychiatric Publishing; May 22, 2013.
2. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
3. Lichstein KL, Durrence HH, Reidel BW, Taylor DJ, Bush AJ. Epidemiology of Sleep: Age, Gender, and Ethnicity. Mahwah, New York: Lawrence Erlbaum Associates Publishers; 2004:202-216.
4. Morphy H, Dunn KM, Lewis M, Boardman HF, Croft PR. Epidemiology of insomnia: a longitudinal study in a UK population. Sleep 2007 Mar;30(3):274-280. [Medline: 17425223]
5. Morin CM, Bélanger L, LeBlanc M, Ivers H, Savard J, Espie CA, et al. The natural history of insomnia: a population-based 3-year longitudinal study. Arch Intern Med 2009 Mar 09;169(5):447-453. [doi: 10.1001/archinternmed.2008.610] [Medline: 19273774]
6. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):125-133. [doi: 10.7326/M15-2175] [Medline: 27136449]
7. Vallières A, Morin CM, Guay B. Sequential combinations of drug and cognitive behavioral therapy for chronic insomnia: an exploratory study. Behav Res Ther 2005 Dec;43(12):1611-1630. [doi: 10.1016/j.brat.2004.11.011] [Medline: 16239154]

8. Morin AK. Strategies for treating chronic insomnia. Am J Manag Care 2006 May;12(8 Suppl):S230-S245 [FREE Full text] [Medline: 16686593]
9. Morin CM, Benca RM. Chronic insomnia. Lancet 2012 Mar 24;379(9821):1129-1141. [doi: 10.1016/S0140-6736(11)60750-2] [Medline: 22265700]
10. Benca RM. Diagnosis and treatment of chronic insomnia: a review. Psychiatr Serv 2005 Mar;56(3):332-343. [doi: 10.1176/appi.ps.56.3.332] [Medline: 15746509]
11. International Narcotics Control Board. World Health Organization. 2010. Report of the International Narcotics Control Board on the Availability of Internationally Controlled Drugs: Ensuring Adequate Access for Medical and Scientific Purposes URL: http://apps.who.int/medicinedocs/en/m/abstract/Js21969en/ [accessed 2019-02-02] [WebCite Cache ID 75tUOiecq]
12. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. J Clin Sleep Med 2008 Oct 15;4(5):487-504 [FREE Full text] [Medline: 18853708]
13. Wilson SJ, Nutt DJ, Alford C, Argyropoulos SV, Baldwin DS, Bateson AN, et al. British Association for Psychopharmacology consensus statement on evidence-based treatment of insomnia, parasomnias and circadian rhythm disorders. J Psychopharmacol 2010 Nov;24(11):1577-1601. [doi: 10.1177/0269881110379307] [Medline: 20813762]
14. Okajima I, Nakamura M, Nishida S, Usui A, Hayashida K, Kanno M, et al. Cognitive behavioural therapy with behavioural analysis for pharmacological treatment-resistant chronic insomnia. Psychiatry Res 2013 Dec 15;210(2):515-521. [doi: 10.1016/j.psychres.2013.06.028] [Medline: 23953170]
15. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia—a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
16. Sato D, Yoshinaga N, Nagai E, Hanaoka H, Sato Y, Shimizu E. Randomised controlled trial on the effect of internet-delivered computerised cognitive-behavioural therapy on patients with insomnia who remain symptomatic following hypnotics: a study protocol. BMJ Open 2018 Dec 30;8(1):e018220 [FREE Full text] [doi: 10.1136/bmjopen-2017-018220] [Medline: 29382675]
17. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. J Psychosom Res 1998 Jul;45(1):5-13. [doi: 10.1016/S0022-3999(97)00298-5] [Medline: 9720850]
18. Doi Y, Minowa M, Uchiyama M, Okawa M, Kim K, Shibui K, et al. Psychometric assessment of subjective sleep quality using the Japanese version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects. Psychiatry Res 2000 Dec 27;97(2-3):165-172. [doi: 10.1016/S0165-1781(00)00232-8] [Medline: 11166088]
19. Yoshinaga N, Niitsu T, Hanaoka H, Sato Y, Ohshima F, Matsuki S, et al. Strategy for treating selective serotonin reuptake inhibitor-resistant social anxiety disorder in the clinical setting: a randomised controlled trial protocol of cognitive behavioural therapy in combination with conventional treatment. BMJ Open 2013;3(2):e002242 [FREE Full text] [doi: 10.1136/bmjopen-2012-002242] [Medline: 23408078]
20. Yoshinaga N, Matsuki S, Niitsu T, Sato Y, Tanaka M, Ibuki H, et al. Cognitive behavioral therapy for patients with social anxiety disorder who remain symptomatic following antidepressant treatment: a randomized, assessor-blinded, controlled trial. Psychother Psychosom 2016;85(4):208-217 [FREE Full text] [doi: 10.1159/000444221] [Medline: 27230862]
21. 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]
22. 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]
23. 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]
24. Seki Y, Nagata S, Shibuya T, Yoshinaga N, Yokoo M, Ibuki H, et al. A feasibility study of the clinical effectiveness and cost-effectiveness of individual cognitive behavioral therapy for panic disorder in a Japanese clinical setting: an uncontrolled pilot study. BMC Res Notes 2016 Oct 07;9(1):458 [FREE Full text] [doi: 10.1186/s13104-016-2262-5] [Medline: 27717407]
25. van Straten A, Emmelkamp J, de Wit J, Lancee J, Andersson G, van Someren EJ, et al. Guided internet-delivered cognitive behavioural treatment for insomnia: a randomized trial. Psychol Med 2014 May;44(7):1521-1532. [doi: 10.1017/S0033291713002249] [Medline: 24001364]
26. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
27. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. J Psychosom Res 2002 Sep;53(3):737-740. [doi: 10.1016/S0022-3999(02)00330-6] [Medline: 12217446]
28. Ye YY, Chen NK, Chen J, Liu J, Lin L, Liu YZ, et al. Internet-based cognitive-behavioural therapy for insomnia (ICBT-i): a meta-analysis of randomised controlled trials. BMJ Open 2016 Dec 30;6(11):e010707 [FREE Full text] [doi: 10.1136/bmjopen-2015-010707] [Medline: 27903557]

29. Feuerstein S, Hodges SE, Keenaghan B, Bessette A, Forselius E, Morgan PT. Computerized cognitive behavioral therapy for insomnia in a community health setting. J Clin Sleep Med 2017 Feb 15;13(2):267-274 [FREE Full text] [doi: 10.5664/jcsm.6460] [Medline: 27784409]
30. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Dec;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
31. Lancee J, van Straten A, Morina N, Kaldo V, Kamphuis JH. Guided online or face-to-face cognitive behavioral treatment for insomnia: a randomized wait-list controlled trial. Sleep 2016 Jan 01;39(1):183-191 [FREE Full text] [doi: 10.5665/sleep.5344] [Medline: 26414893]
32. Yamadera W, Sato M, Harada D, Iwashita M, Aoki R, Obuchi K, et al. Comparisons of short-term efficacy between individual and group cognitive behavioral therapy for primary insomnia. Sleep Biol Rhythms 2013 Jul;11(3):176-184 [FREE Full text] [doi: 10.1111/sbr.12019] [Medline: 24098091]
33. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017 Feb 15;13(2):307-349 [FREE Full text] [doi: 10.5664/jcsm.6470] [Medline: 27998379]
34. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
35. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]
36. Buysse DJ, Thompson W, Scott J, Franzen PL, Germain A, Hall M, et al. Daytime symptoms in primary insomnia: a prospective analysis using ecological momentary assessment. Sleep Med 2007 Apr;8(3):198-208 [FREE Full text] [doi: 10.1016/j.sleep.2006.10.006] [Medline: 17368098]

Abbreviations

ACP: American College of Physicians
ANCOVA: analysis of covariance
CBT: cognitive behavioral therapy
CES-D: Center for Epidemiologic Studies Depression Scale
CONSORT: Consolidated Standards of Reporting Trials
EQ-5D: EuroQol-5D
HADS: Hospital Anxiety and Depression Scale
ICBT: internet-delivered computerized cognitive behavioral therapy
NA: number of awakenings
PSQI: Pittsburgh Sleep Quality Index
QALY: quality-adjusted life year
QOL: quality of life
RCT: randomized controlled trial
SE: sleep efficiency
SOL: sleep onset latency
TST: total sleep time
UC: usual care

Edited by G Eysenbach; submitted 04.11.18; peer-reviewed by Y Minabe, N Siriwardena; comments to author 05.12.18; revised version received 19.12.18; accepted 21.12.18; published 11.04.19.

Please cite as:
Sato D, Yoshinaga N, Nagai E, Nagai K, Shimizu E
Effectiveness of Internet-Delivered Computerized Cognitive Behavioral Therapy for Patients With Insomnia Who Remain Symptomatic Following Pharmacotherapy: Randomized Controlled Exploratory Trial
J Med Internet Res 2019;21(4):e12686
URL: http://www.jmir.org/2019/4/e12686/
doi:10.2196/12686
PMID:30973344

Original Paper

Web-Based Parent Training Intervention With Telephone Coaching for Disruptive Behavior in 4-Year-Old Children in Real-World Practice: Implementation Study

Terja Ristkari[1*], RN, MSN; Marjo Kurki[1*], PhD; Auli Suominen[1], MSci; Sonja Gilbert[1], PhD; Atte Sinokki[1], BEng; Malin Kinnunen[1], RN; Jukka Huttunen[1], BM; Patrick McGrath[2,3,4], OC, PhD, FRSC, FCAHS; Andre Sourander[1*], MD, PhD

[1]Department of Child Psychiatry, Turku University Central Hospital, University of Turku, Turku, Finland
[2]Centre for Research in Family Health, Izaak Walton Killam Health Centre, Halifax, NS, Canada
[3]Strongest Families Institute, Halifax, NS, Canada
[4]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada
[*]these authors contributed equally

Corresponding Author:
Terja Ristkari, RN, MSN
Department of Child Psychiatry
Turku University Central Hospital
University of Turku
Lemminkäisenkatu 3, Teutori 3rd Floor
Turku, 20014
Finland
Phone: 358 504643141
Email: terjaris@utu.fi

Abstract

Background: Parent training is the most effective approach to the psychosocial treatment of disruptive behavioral problems in childhood. However, no studies exist on how well Web-based training programs work when they make the transition from the research setting to implementation in primary health care.

Objective: The study aimed to examine how the randomized controlled trial (RCT) and implementation study groups of the Strongest Families Smart Website (SFSW) intervention differed in child psychopathology, family demographics and treatment-related factors, such as therapeutic alliance and parents' satisfaction rates. The intervention was conducted in the pediatric primary health care in Finland.

Methods: The study focused on 232 parents who had taken part in the SFSW intervention, which formed part of a 2-arm RCT study, and 882 families that would participate in the subsequent SFSW implementation study group. Both groups comprised parents whose children displayed high levels of parent-reported disruptive behavioral problems when they were screened in child health clinics at 4 years of age. Parents in both groups were provided with the SFSW intervention, which consisted of a Web-based training program with 11 weekly themes and associated telephone sessions.

Results: Demographic factors or duration of behavioral problems did not differ statistically or clinically between the RCT and implementation groups. Overall, 42.0% (362/862) of children in the implementation group and 35.4% (80/226) in the RCT intervention group had suffered from behavioral difficulties more than 1 year before the screening phase (χ_1^2=3.2; P=.07). The mean duration of telephone coaching calls was very similar in the implementation and RCT intervention groups, that is, 38 and 37 min per call, respectively ($t_{279.5}$=0.26; P=.79). The total time spent on the website of the program was 451 min in the implementation group and 431 min in the RCT intervention group ($t_{318.8}$=1.38; P=.17). In the RCT intervention group, 52 of the 232 participants (22.4%) discontinued the program before the tenth week, whereas in the implementation group, 109 of the 882 participants (12.4%; odds ratio 2.05, 95% CI 1.4-3.0; P<.001) discontinued. Parents in both the implementation (77.1% to 98.5%, 498/742 to 731/742, respectively) and the RCT (64.8% to 98.2%, N=105/162- to 159/162, respectively) groups reported qualitatively similar and high level of posttreatment satisfaction rates in improved parenting skills, expectations, and stress relief. Parents in both groups reported a high level of satisfaction in skills and professionalism of the telephone coaches.

Conclusions: The implementation of population-based screening of Web-based parent training intervention with telephone coaching resulted in good feasibility, fidelity, accessibility, and similar satisfaction level post treatment when compared with intervention in RCT research setting. The discontinuation of treatment in the implementation group was exceptionally low.

(J Med Internet Res 2019;21(4):e11446) doi:10.2196/11446

KEYWORDS

child mental health; early intervention; parenting education; disruptive behavior; preschool children

Introduction

Background

Parent training is the most effective approach to the psychosocial treatment of disruptive behavioral problems in childhood, and there is mounting evidence from randomized controlled trials (RCTs) that such initiatives reduce problems and improve parenting skills [1-3]. It has been proposed that parent training should be provided as primary promotion and prevention in primary care settings [4]. However, only a small proportion of families with these problems benefit from evidence-based treatment programs [5]. The barriers to receiving parent training include the lack of trained staff that can provide interventions; the stigma related to receiving mental health treatment; and the difficulties of accessing and engaging in treatment in terms of costs, time, and location [6,7].

Technology-based parent training programs can offer many benefits over traditional interventions, such as higher fidelity, greater accessibility, convenience, and reduced time and costs [8-11]. Technology-based parent training is not a new innovation because, as early as 1988, Webster-Stratton et al [12] tested videotapes as the primary delivery method for a parent training intervention. Today, technology and internet are integral parts of people’s lives. For example, in 2017, 88% of Finnish people used the internet, including almost every adult under the age of 55 years, and 77% had smartphones, which is one of the easiest ways to access the internet [13]. Accordingly, in the field of parent training as primary care of children’s conduct problems, very recent research has shifted to focus on Web-based training programs. Studies have shown promising efficacy of Web-based interventions in improving child behavior [8,14]. Especially, interactive Web-based programs have been found to be more effective than noninteractive programs [14]. Such Web-based interactive parent training programs could overcome many barriers associated with most traditional programs in the implementation phase, especially concerning the consistency with the original evidence-based intervention [10].

We previously reported the 12-month follow-up study of the first RCT to provide an interactive Web-based parent training program with supplementary weekly phone coaching, the Strongest Families Smart Website (SFSW), using a population-based screening procedure [15]. The 12-month follow-up study showed that the intervention resulted in significant reductions in the level of disruptive behavior problems among 4-year-old children and improved parenting skills. The treatment outcomes remained significant at 24-month follow-up when the intervention group was compared with the control group [16]. However, no research exists about how this model works when the program is implemented in primary health care settings.

Regarding psychosocial interventions, there are 2 different aspects when converting these interventions from the research environment to the *real world*: dissemination and implementation of an innovation in clinical practice. Dissemination refers to how knowledge of the new practices is extended actively and passively, whereas implementation refers to the action of accommodating new practices into real treatment environments [17]. *Implementation gap* refers to the difference between our knowledge of *what works* and *how it works* and the application of this knowledge in real-world practice. Research on the implementation of digital mental health interventions into routine care is scarce and studies on such interventions for preschool children are almost nonexistent. A consistent finding is that the vast majority of children with psychiatric problems early in life have unmet needs as only a minority of children with problems are referred and there are substantial delays in contacting specialist services [18]. The great majority of adult psychiatric disorders begin in childhood or adolescence [19]. Therefore, implementation of effective evidence-based treatments in real life plays a key role in child mental health service research [20].

Previous implementation studies have emphasized that certain core implementation components and quality assurance measures ought to be included in the implementation strategy [21]. Typically, the effects seen in the RCT settings decline during implementation [8]. These quality measures are mostly related to the practitioner (referred to as coaches herein) of the program [22,23]. We ensured strict adherence to the same study protocol procedures as in the RCT study by accounting for certain quality measures as explained in the implementation plan (see Methods).

Objectives

In this study, we implemented our aforementioned RCT study in primary health care settings to see how it would work in the real world. The first aim of this study was to compare certain child and family characteristics of the RCT intervention group with those of the first 882 families who received treatment during the primary care implementation phase. Both groups were based on population-based screening of 4-year-old children with high levels of disruptive behavior [24]. We were particularly interested in finding out whether the level of child psychopathology, duration of problems, and impairment levels were similar when the RCT intervention group and the implementation group were compared. The second aim was to examine the differences between the 2 study groups in certain elements of the program, such as the time the parents spent reading psychoeducational material and completing skill exercises. The third aim was to compare the satisfaction levels in the 2 groups, namely, how the program affected their parenting skills, parental stress, and satisfaction with coaching.

We hypothesized that the screening procedure would also work in the implementation phase and no major differences would be found in the demographic factors or child psychopathology profiles when compared with the RCT study population. We also anticipated no major differences in the content of the program or parental satisfaction levels between the 2 groups.

Methods

Context of the Study

The study took place in Finland, a Nordic welfare state that provides its residents with public health services. Finnish child health clinics provide annual checkups that offer universal health care and are attended by 99.6% of children [25]. The clinics try to identify problems that affect families with small children at an early stage and arrange for them to receive appropriate help. All parents are invited to bring their child to the child health clinic checkups about 15 times from birth to the age of 6 years. When the children attend the checkup at 4 years of age, they have reached a suitable stage in their development for identification of disruptive symptoms and provision of early support for families.

Study Population

This study compared the intervention group from the RCT study with the first 882 families to receive parent training during the implementation phase. In both groups, all children attending child health clinics checkups at 4 years were screened. Figure 1 shows the flow chart of the study.

The complete RCT study protocol has previously been described in detail [26]. In summary, the study design was an RCT with 2 parallel groups that were stratified by sex with 1:1 individual allocation. From 2011 to 2013, 5 municipalities in Southwest Finland joined the study. After the initial screening of 4656 children, 15.68 % (730/4656) 4-year-old children with high levels of disruptive behavioral problems were identified, and 464 (10.0%) parents were eligible to take part in the study and agreed. These parents were randomized into the SFSW intervention group (N=232) or an education control (N=232). The participants randomized to the intervention received an 11-session, internet-based parenting program that focused on skills for strengthening their parent-child relationships, together with a series of weekly telephone coaching sessions.

The implementation study group comprised the first 882 families to participate in the study following the success of the RCT. Initially, 12,780 children were screened, 1663 (13.01%) children with high levels of disruptive behavioral problems were identified, and 882 (6.9%) of the parents of 4-year-old children were eligible and agreed to take part in the study.

Implementation Plan

After an RCT study, when converting an evidence-based intervention into practice in a real-world setting, it is essential to tailor discrete multicomponent strategies for the implementation [27]. Here, we identified the core implementation components of our implementation strategy. The following act as implementation drivers [21]:

Recruitment, Staff Selection and Training

All coaches were professionals in health care (eg, public health nurses and nurses) and social services (children's services). In our model, conduction and coaching of the intervention were centralized in the Research Centre for Child Psychiatry at the University of Turku. Each coach was trained on the protocol of the digital program by experienced Strongest Families clinicians. The training consisted of a theoretical background (such as preventive mental health and conduct behavior problems) and rehearsal phone calls. After the basic training, the coaches started the program by recruiting families. Then, they progressed to the closely monitored coaching calls (see below) with the actual families.

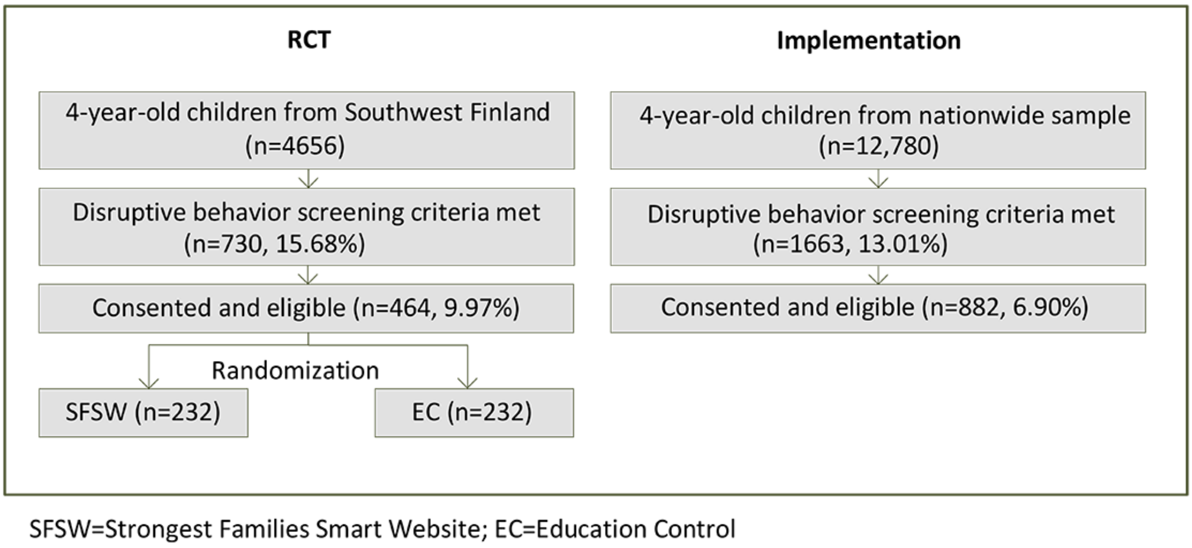
Before the utilization of the treatment model (screening and parent training intervention), a half-day training session was organized for the key actors in primary health care (such as public health nurses and doctors) to introduce the background of the program and treatment model and to make announcements and distribute other material.

In primary health care, the child health nurses were kept up to date on the progress of their clients within the Strongest Families program to ensure that they adhered to the parent training intervention model. This included summaries of the screening questionnaires complete with clinical outcome recommendations, updates on families consenting to the program, and a brief summary of the treatment program outcomes. In addition, monthly reports of questionnaire participation rates and numbers of completed treatments and dropouts were sent to child health clinic managers and chief doctors.

Ongoing Supervision and Staff Performance Evaluation

We conducted systematic, weekly supervision meetings with the individual coaches and group case conferences, where all coaches reviewed and discussed the families they were coaching. The more experienced coaches acted as supervisors. After each telephone coaching call, the coach assessed his or her own performance on a scale from 4 to 10. If self-assessment was greater than or equal to 6, the supervisor received a message from the digital platform and subsequently discussed the issue with the coach. About 10% of the recorded coaching calls were audited by the coach supervisor and evaluated for competency, with additional training and monitoring of future calls, when indicated.

Figure 1. Randomized controlled trial (RCT) intervention and implementation flow charts.



Decision Supporting and Administration

The treatment model was introduced to the stakeholders of children and family services who decided the budget for the region or municipality. All parties signed the jointly funded annual research agreement on the treatment model implementation. Regular contact was maintained with the key actors of the region. The stakeholders received user-friendly monthly progress reports including the number of screened and enrolled families, etc. To integrate the model as a part of primary health care, regular training was conducted and information about the progress of the implementation study was distributed. Moreover, we involved local and national media to increase public awareness of our treatment model and more generally child mental health issues. Centralization of delivery of the digital parent training intervention and the development of the digital platform were the core of the facilitative administration. These strategies strengthened the implementation fidelity, which refers to the degree of adherence to implementation strategy and adaption process into local practices [28].

Of the original municipalities in the RCT study, the City of Turku and Town of Naantali, continued in the program when the implementation study began. Within the first year of the implementation, several municipalities and 2 provinces throughout Finland joined the study. The new treatment model had received publicity in media, and the RCT findings had been presented at national professional and scientific meetings.

The 882 families in the implementation phase came from throughout Finland, specifically the cities of Tampere, Espoo, Kouvola, Kuopio, and Turku; the towns of Lahti, Hyvinkää, Naantali, and Parainen; the municipalities of Kittilä, Mäntsälä, and Tuusula; and 3 provinces, South Karelia Social and Health Care District, Eksote, and Kainuu Social Welfare and Health Care Joint Authority.

A basic study plan was drafted by the research team for the implementation study, and this was introduced to senior child health clinic and health care officials from each of the health care districts. No major revisions were needed to accommodate regional needs.

A significant difference between the RCT and implementation study was that there was no control group, and the first 882 families that met the inclusion criteria were offered the chance to participate in the program. The child health nurses and their supervisors were key to introducing the early intervention program to the families.

Population-Based Screening Procedure and Recruitment

The screening procedure was similar in the RCT study and implementation phase, and it was integrated into the standard 4-year-old child health checkup visit. First, the supervising child health clinic staff and nurses in the participating areas were informed of the screening and intervention. Then, a 2-hour training session was held in each municipality to introduce the nurses to the background of the program and study plan and what they needed to do. We collected the data for all parents with children turning 4 years after January 2015 from the population register, and letters were mailed to parents in the middle of the month before the child’s fourth birthday. The letters included a short newsletter about the program, the form including questions of demographic information of parents and child, the Strengths and Difficulties Questionnaire (SDQ) [29-31], and instructions for bringing the filled form to the annual checkup at the child health clinic. After the checkup, the nurses mailed the completed form to the research staff, and the answers were entered into the study database and scored with Access software (Microsoft Corp, Redmond, WA, USA). All the questionnaires were then mailed back to the respective child health clinics along with a separate SDQ score sheet, where they were reviewed together with the parents at their next visit to the child health clinic.

The inclusion and exclusion criteria for the implementation study were identical to the RCT trial. Parents were included in the study if their child scored 5 or more on the conduct subscale of the SDQ questionnaire and their answers to the SDQ impact supplement indicated that their child had behavioral problems. Parents were excluded if the child did not live with them because they were under the care of child protection services because of child custody, abuse, or neglect issues; the child did not speak in full sentences; or the parents had participated, or were participating, in other parent training or behavioral treatment. Children were also excluded if they met the diagnostic criteria for autism or a pervasive development disorder, Down syndrome, fetal alcohol syndrome, mental retardation, or a genetic diagnosis that would lead to mental retardation.

If the parents appeared to qualify for the study, they received a phone call from the program's recruitment staff. They provided a brief introduction to the parent training program and reviewed the exclusion criteria. If the parents wanted to participate, they were directed to the study website to provide their formal consent. A form summarizing the recruitment call was mailed to the referring child health clinic to advise them of the outcome, as the municipalities were obliged to offer alternative treatment options for families not willing to participate in the parent training.

When the parents first logged into the study website, they were presented with the Web-based consent form for the study and a nonbinding program agreement, where the principles for successfully completing the program were outlined. These principles included adhering to the mutually agreed call schedule. The goal of the agreement was to ensure that the parents starting the program were reasonably committed to completing the training, which lasted for 3-4 months.

Parent Training Intervention

The SFSW intervention consisted of 2 components: the interactive website and the telephone coaching. The parent training program was delivered in a Web-based environment that the parent and coach both had access to, they had scheduled weekly phone calls, and they could also send additional messages via the website. No physical visits or face-to-face communication was conducted at any point during the program.

The program consisted of 11 weekly themes with associated phone coaching, and the program started with an introductory call from the coach after the parent had completed the baseline surveys. This provided an overview of the general objectives of the program and explained the learning methods and need for cooperation. It also set the initial goals for the program and introduced the first of the 11 weekly themes, which was called *noticethegood*. The goals were set in response to the problem behaviors that the child exhibited and that the program aimed to alleviate by teaching the parent how to use positive parenting skills and problem-solving abilities.

After the introduction, the parent was directed to the material for the first Web-based session, completed the associated skill training, and received the next coaching call a week later. The coaches used different verbal techniques during the phone calls, such as discussing different parenting models and role play as well as encouraging parents to practice every day. By using attributional questions, the coach could motivate parents to reflect on their own behavior as well as their child's. The website tracked the parent's activities on the site, and if they had not logged on for 2 days, the sites sent them a reminder, suggesting that they log in and complete the activities. The coach was notified if the parents had not been active on the website within 4 days of the last call. They then contacted the parent, encouraged to proceed with the program, and, if necessary, rescheduled the next call. The structured coach interaction with each client and monitoring are the key elements that distinguish coach-guided parent training from other forms of parent training, such as group-based, self-help, or email-assisted programs.

The content of each Web-based session was divided into the introduction, session content, video exercises, troubleshooting, review, and practical application of the new skills. Each section contained interactive and multimedia components, such as exercises and video clips. The parent was encouraged to complete the session by the next phone call. Some sessions also included supplementary material, which was emailed separately. The Web-based sessions followed the chronological order as shown in Table 1.

The sessions were conceptually divided into 3 sections: (1) basic positive parenting skills, (2) practical parenting skills, and (3) reinforcing acquired skills and sustained positive parenting. The aim of the basic positive parenting skills was to refocus the parent's attention from the child's problem behavior to noticing the child's success in everyday life, as reflected in the name of the first session, notice the good. It also aimed to change the parent's reactions to the child's behavior from a negative to a positive response. Practical parenting skills focused on applying the basic skills in everyday situations, planning ahead with regard to daily activities, and using the supporting methods to reinforce positive behavior. These included sticker charts, when and then statements, and using time-out as a way to help the child and parent to regulate their emotions.

During the program, the parent learnt to solve problem situations using positive and practical skills and to understand their child better from a developmental and emotional perspective. The final sessions focused on reinforcing how they had applied the skills they had learnt, independent of the coaching support, and how they could sustain those skills beyond the active program.

Table 1. Structure of the web-based element of the Strongest Families Smart Website training program.

Session	Key training elements	Parental goals	Supplementary material
Notice the good	Positive and active parenting	Boosts child's self-esteem; boosts parent's self-esteem; and changes the parent's view of their child	—[a]
Spread attention around	Positive, impartial parenting	Strengthens child's empathy skills	—
Ignore whining and complaining	Positive, self-controlled parenting	Teaches parents self-regulation	—
Prepare for changes	Positive, proactive parenting	Reinforces good daily routines	—
Plan ahead at home	Positive, proactive parenting	Boosts child's and parent's self-esteem and involves the child in planning	—
Chart and stickers	Positive, active parenting	Involves the child in planning and reinforces good daily routines	Sticker chart and stickers
Plan ahead outside home	Positive, proactive parenting	Boosts child's and parent's self-esteem and involves the child in planning	—
Working with daycare	Positive cooperation and /communication between parent and day care	Helps child to manage and succeed	Daily report card
Time-out	Positive, self-controlled parenting	Teaches self-regulation and consistency	Digital timer
Problem solving, revision, and future application of skills	Positive daily parenting in future	Teaches parents skills to support child development and prepares for upcoming challenges	—
Booster	Skills review	Reminds parents of positive proactive parenting skills	Skills review chart

[a] Not applicable.

Measurements Used in This Study

Demographic information of family and parents was obtained at the screening phase, and the variables that were included were information about the sex of the child, family structure, and the parents' birth year and education and whether they had been unemployed. Child variables based on SDQ were collected from the form filled by parents at the screening phase.

Psychopathology was screened using the Finnish version of the SDQ for the parents of children aged 2 to 4 years [30], which is widely used as a screening and research tool and for clinical assessment and outcome evaluation. The SDQ consists of 25 items covering both positive and negative behaviors. These are divided into the 5 subscales of symptoms: emotional problems, conduct problems, hyperactivity or inattention, peer relationship problems, and prosocial behavior. Each subscale consists of 5 questions, and each item is rated on a scale with 3 possible answers: never (0 points), somewhat true (1 point), and certainly true (2 points). The SDQ impact supplement was also used, before inclusion in the program, to determine whether the child had a problem. If the parent said that they felt their child did have a problem, they were asked how chronic the problem was and about the distress, social impairment, and burden that the problem caused to others. The perceived difficulties were assessed with a single question: "Overall, do you think that your child has difficulties in one or more of the following areas: emotions, behavior or being able to get on with other people?" The alternatives were no, minor difficulties, definite difficulties, and severe difficulties.

Each participant's time on the website was downloaded using appropriate time-out values, including the percentage of primary screens that the participants observed. The duration of therapeutic calls was downloaded and summarized at the end of the treatment. Data on treatment-related factors, program satisfaction, and therapeutic alliance questions were collected and provided to the research team at the end of treatment. The treatment-related factors covered 3 domains. The first was where the coaching took place (at home, at work, or in another place) and how often the coaching calls were received (every week/almost every week, a couple of times/once during the program, or never). The second was where the materials were read (at home, at work, or in another place), how often they were read (daily/almost daily, a couple of times a week/once a week, or never), and whom they read the material with (alone or with the child's other parent). The third was the parents' satisfaction with the program and therapeutic alliance questionnaire that explored their relationship with the program coach. These used a 5-point Likert scale (strongly disagree, disagree, not agree or disagree, agree, and strongly agree) and contained 13 propositions and 3 sections, covering the family program (the program matched my expectations, I would recommend the program to my friends, if they were in need of similar help, in case I would need help in future, I would enroll to the program again, generally speaking, how content have you been with the Strongest Family program?), the effect of the program on their parenting skills (I have learned skills, which have been helpful to me as a parent, I trust more my abilities to act as a parent, my relationship with my child has improved, my stress levels have been relieved), and the family coach (the coach respected my views on parenting, the coach was professional, the coach encouraged problem solving, I could form a successful working relationship with the coach).

Ethical Approval

Ethical approval was received from the research ethics boards of the Hospital District of Southwest Finland and our Canadian program partners, the IWK Health Centre, Halifax, Nova Scotia, for the RCT study. We received ethical approval from the University of Turku for the implementation study.

Data Analysis

To explore the differences between the implementation and the RCT intervention, we used the Pearson Chi-square test for categorical variables or the Fisher exact test, if the assumption of expected counts was violated. The assumption of equal variances and residuals normality was tested for continuous variables. If the assumptions were valid, the independent group *t* test was used. If normality assumptions failed, the *t* test was repeated after logarithmic transformation. If the log-transformed variable was not reaching assumptions, the Mann-Whitney U test with normal approximation for large samples (Z test) was applied. The effect sizes were estimated by the Cramér V for Chi-square test and Fisher exact test and by the Cohen *d* for *t* test and Z test.

The odds ratios (ORs) with 95% CI were estimated using logistic regression to examine the associations between discontinuation in the implementation and RCT groups. The statistical significance was classified by a 2-way *P* value of <.05. The statistical analyses were conducted using SAS statistical software version 9.4 (SAS Institute Inc, Cary, NC, USA).

Results

The RCT group (N=464 including both intervention and control group) was screened from a population of 4656 children who were 4 years old (9.97%), whereas the implementation group including 882 children was screened from 12,780 children who were 4 years old (6.90%; χ_1^2=45.0; *P*<.001). Of those 232 who were in the RCT intervention group, 52 (22.4%) discontinued the program before the tenth week, whereas the respective figure for those 882 in the implementation group was 109 (12.4%). The logistic regression indicated that parents in the implementation group were twice as likely to complete the program than in the RCT intervention group (OR 2.05, 95% CI 1.4-3.0; *P*<.001).

Family Demographics and Child Psychopathology

Tables 2 and 3 shows the family demographic and child psychiatric information for the 882 children (553/882, 62.7% boys) in the implementation group and the 232 children (142/882, 61.2% boys) in the RCT intervention group. No differences were observed in family structure, parental age, or education level between the 2 groups. When the SDQ psychopathology of the children was compared, the groups had very similar scores. However, the implementation group had more moderate or severe difficulties (53.1%, 467/880; *P*=.02; effect size=−.071) and more severe impairment based on SDQ impact score (mean 1.1, SD 1.4; *P*=.006; effect size=0.214) than the RCT intervention group (44.4%, 103/232; mean 0.8, SD 1.2. It was notable that 42.0% (362/862) of the children in the implementation group and 35.4% (80/226) of the children in the RCT intervention had suffered from behavioral difficulties for more than a year before they were assessed.

Treatment-Related Factors

As shown in Tables 4 and 5, most of the treatment variables were very similar in both groups. The telephone coaching calls were mostly received at home. Parents who took part in the RCT intervention phase were more likely to receive calls and read the material at work than the parents in the implementation group. Overall, the implementation group received an average of 11 calls (SD 2.2), whereas the RCT group received an average of 10 calls (SD 3.3; Z=3.41; *P*<.001) over the course of the program. The mean duration of the calls was very similar in both groups, lasting 37 to 38 min per call. The time spent on the website during the whole 11-week program was a mean of 451 (SD 174) min in the implementation group and 431 (SD 207) min in the RCT group ($t_{318.8}$=1.83; *P*=.17).

Table 2. Family and child factors in the implementation (N=882) and randomized controlled trial intervention (N=232) groups. Statistically significant values (P<.05) are shown in italics.

Family factors	Implementation, n (%)	Randomized controlled trial intervention, n (%)	Chi-square test (df)	*P* value	Effect size[a]
Family and parent variables					
Sex of the child			**0.2 (1)**	**.68**	**0.013**
Male	553 (62.7)	142 (61.2)			
Female	329 (37.3)	90 (38.8)			
Family structure			**0.3 (1)**	**.58**	**−.017**
Two biological parents	720 (81.9)	192 (83.5)			
Other	159 (18.1)	38 (16.5)			
Maternal age at childbirth			**0.4 (1)**	**.55**	**0.018**
Up to 24 years	112 (12.8)	26 (11.3)			
25 plus years	764 (87.2)	204 (88.7)			
Paternal age at childbirth			**2.4 (1)**	**.12**	**0.048**
Up to 24 years	69 (8.2)	11 (5.1)			
25 plus years	778 (91.9)	207 (95.0)			
Maternal education			**1.0 (1)**	**.31**	**−.031**
Comprehensive school or lower or secondary education	339 (38.9)	98 (42.6)			
College/university degree	532 (61.1)	132 (57.4)			
Paternal education			**0.1 (1)**	**.71**	**0.012**
Comprehensive school or lower or secondary education	445 (54.7)	115 (53.2)			
College/university degree	369 (45.3)	101 (46.8)			
Child variables based on Strengths and Difficulties Questionnaire					
Difficulties			**5.5 (1)**	***.02***	**−.071**
Minor difficulties	413 (46.9)	129 (55.6)			
Moderate or severe difficulties	467 (53.1)	103 (44.4)			
Length of difficulties			**3.2 (1)**	**.07**	**− .055**
Less than 1 year	500 (58.0)	146 (64.6)			
1 year or more	362 (42.0)	80 (35.4)			

[a] The effect size is measured by Cramér V.

Table 3. Family and child factors in the implementation (N=882) and randomized controlled trial intervention (N=232) groups. Statistically significant values (P<.05) are shown in italics.

Family factors	Implementation, mean (SD)	Randomized controlled trial intervention, mean (SD)	Z-test	*t* test (df)	*P* value	Effect size[a]
Total score	15.5 (4.6)	14.9 (4.4)	—	1.76 (1108)	.08[a]	0.139[b]
Emotional score	1.9 (1.8)	1.7 (1.5)	−1.1	—	.29[b]	0.128
Peer problems score	2.4 (1.7)	2.3 (1.6)	—	1.16 (1109)	.25[a]	0.081
Prosocial behavior	5.9 (1.9)	6.0 (1.8)	—	−0.80 (1112)	.43	−.059
Hyperactive score	4.9 (2.5)	4.7 (2.3)	—	1.14 (1111)	.26	0.086
Conduct score	6.2 (1.3)	6.2 (1.3)	—	0.49 (1112)	.62[a]	0.042
Impact score	1.1 (1.4)	0.8 (1.2)	−2.8	—	*.006*[b]	0.214

[a] The *P* value after logarithmic transformation by *t* test.

[b] The *P* value was determined by the Mann-Whitney U test with normal approximation for large samples.

Table 4. Treatment-related categorical factors in the implementation (N=882) and randomized controlled trial intervention (N=232) groups, excluding 21 parents who did not have any coaching calls. Statistically significant values (*P*<.05) are shown in italics.

Treatment-related factors	Implementation, n (%)	Randomized controlled trial intervention, n (%)	Chi-square test (df)	Fisher exact test	*P* value	Effect size[a]
Where calls were received?						
At home			**2.6 (2)**	**—[b]**	**.27**	**.055**
Every week/almost every week	563 (76.2)	125 (82.2)				
Couple times/once during the program	148 (20.0)	23 (15.1)				
Never	28 (3.8)	4 (2.6)				
At work			**5.8 (2)**	**—**	**.054**	**.082**
Every week/almost every week	118 (16.0)	31 (23.3)				
Couple times/once during the program	146 (19.8)	30 (22.6)				
Never	475 (64.3)	72 (54.1)				
Other place than home or work			**3.1 (2)**	**—**	**.21**	**.60**
Every week/almost every week	54 (7.3)	7 (5.7)				
Couple times/once during the program	306 (41.4)	42 (34.4)				
Never	379 (51.3)	73 (59.8)				
Where materials were read?						
At home			**0.5 (2)**	**—**	**.79**	**.023**
Daily/almost daily	50 (6.8)	11 (7.0)				
Couple times a week/once a week	680 (92.0)	144 (91.1)				
Never	9 (1.2)	3 (1.9)				
At work			**—**	**0.0001**	***.002***	**.117**
Daily/almost daily	2 (0.3)	1 (0.7)				
Couple times a week/once a week	193 (26.1)	57 (39.9)				
Never	544 (73.6)	85 (59.4)				
Other place than home or work			**4.2 (2)**	**—**	**.12**	**.070**
Daily/almost daily	7 (1.0)	1 (0.8)				
Couple times a week/once a week	152 (20.6)	16 (12.8)				
Never	580 (78.5)	108 (86.4)				
Whom were materials read with?						
Alone			**0.8 (2)**	**—**	**.77**	**.024**
Daily/almost daily	41 (5.6)	10 (6.4)				
Couple times a week/once a week	661 (89.5)	141 (89.8)				
Never	37 (5.0)	6 (3.8)				
With the other parent of the child			**0.7 (2)**	**—**	**.69**	**.029**
Daily/almost daily	10 (1.4)	2 (1.4)				
Couple times a week/once a week	423 (57.2)	78 (53.4)				
Never	306 (41.4)	66 (45.2)				
Completed part of the program			**24.3 (3)**	**—**	***<.001***	**.148**
Only introduction part	16 (1.8)	14 (6.0)				
Basic skills (weeks 1-4)	51 (5.8)	18 (7.8)				
Functional skills (weeks 5-9)	45 (5.1)	24 (10.3)				
Revision and future application of skills (weeks 10-11)	770 (87.3)	176 (75.9)				

[a] The effect size is measured by Cramér V.

[b] Not applicable.

Table 5. Treatment-related continuous factors in implementation (N=882) and randomized controlled trial intervention (N=232) groups, excluding 21 parents who did not have any coaching calls. Statistically significant values (*P*<.05) are shown in italics.

Treatment-related factors	Implementation		Randomized controlled trial intervention		Z-test	*t* test (df)	*P* value[a]	Effect size[b]
	Mean (SD)	Range	Mean (SD)	Range				
Number of calls	11 (2.2)	2-14	10 (3.3)	1-14	3.41	—[c]	*<.001*	.146
Duration of call (min)	38 (10.3)	3-102	37 (13.5)	2-82	—	0.26 (279.5)	.79	.021
Duration of sign-ins (by staying on the website, min)	451 (174)	72-1330	431 (207)	29-1362	—	1.38 (318.8)	.17	.107

[a] The *P* value was determined by Mann-Whitney U test with normal approximation for large samples.

[b] The effect size is measured by Cohen *d*.

[c] Not applicable.

Parents' Satisfaction

As Table 6 shows, the participants in both groups reported high levels of satisfaction with how the program had improved their parental skills. The satisfaction questionnaire was filled by 744 of 882 participants (84.4%) in the implementation and 162 of 232 participants (69.8%) in the RCT intervention phase. In the implementation group, 83.4% (619/742) to 98.5% (731/742) reported high satisfaction with various aspects of the program, and in the RCT group, 82.7% (134/162) to 98.2% (159/162) were very highly satisfied. When they were asked about the overall impact of the program, 88.4% (658/744) of the parents in the implementation groups said it had matched their expectations and 94.0% (699/744) said it had matched their needs. The respective figures for the RCT intervention group were 82.7% (134/162) and 93.8% (152/162). In addition, 77.1% (498/742) of implementation group parents and 64.8% (105/162) of the RCT parents reported that the program had reduced their stress. Finally, both groups reported very high levels of satisfaction (96.8% [719/743] and 100% [162/162], respectively) with the skills and professionalism of the telephone coaches and their relationships with them. We conducted a sensitivity analysis on the satisfaction-related factors, where we excluded 2 municipalities from the implementation sample that participated in the RCT study (altogether 159 cases). The results were not qualitatively different from the tests on the complete data.

Table 6. Satisfaction-related factors in implementation (N=882) and randomized controlled trial intervention (N=232) groups, excluding 21 parents who did not have any coaching calls. Statistically significant values (*P*<.05) are shown in italics.

Satisfaction-related factors[a]	Implementation, n (%)	Randomized controlled trial intervention, n (%)	Chi-square test (df)	Fisher exact test	*P* value	Effect size[b]
Overall impact: family program						
The program matched my expectations			**3.96 (1)**	—[c]	***.047***	**−.066**
Disagree/neutral	86 (11.6)	28 (17.3)				
Agree	658 (88.4)	134 (82.7)				
The program met my needs			**0.004 (1)**	—	**.95**	**−.002**
Disagree/neutral	45 (6.1)	10 (6.2)				
Agree	699 (94.0)	152 (93.8)				
I would recommend the program to my friends, if they were in need of similar help			**1.01 (1)**	—	**.31**	**.033**
Disagree/neutral	25 (3.4)	3 (1.9)				
Agree	719 (96.6)	159 (98.2)				
In case I would need help in the future, I would enroll to the program again			**0.16 (1)**	—	**.69**	**−.013**
Disagree/neutral	97 (13.0)	23 (14.2)				
Agree	647 (87.0)	139 (85.8)				
Generally speaking, how content have you been with the Strongest Family program?			**—**	**0.20**	**.18**	**.020**
Disagree/neutral	20 (2.7)	3 (1.9)				
Agree	724 (97.3)	159 (98.2)				
Overall impact: effects on parenting skills						
I have learned skills, which have been helpful to me as a parent			**—**	**0.09**	**.18**	**−.047**
Disagree/neutral	11 (1.5)	5 (3.1)				
Agree	731 (98.5)	157 (96.9)				
I trust more in my abilities to act as a parent			**0.39 (1)**	—	**.53**	**.021**
Disagree/neutral	86 (11.6)	16 (9.9)				
Agree	656 (88.4)	146 (90.1)				
My relationship with my child has improved			**0.03 (1)**	—	**.87**	**.006**
Disagree/neutral	123 (16.6)	26 (16.1)				
Agree	619 (83.4)	136 (84.0)				
My stress levels have been relieved			**0.31 (1)**	—	**.57**	**−.019**
Disagree/neutral	244 (32.9)	57 (35.2)				
Agree	498 (77.1)	105 (64.8)				
Direct impact: family coach						
The coach respected my views on parenting			**—**	**0.04**	**.09**	**.063**
Disagree/neutral	16 (2.2)	0 (0)				
Agree	726 (97.8)	162 (100.0)				
The coach was professional			**—**	**0.16**	**.40**	**.034**
Disagree/neutral	19 (2.6)	2 (1.2)				
Agree	723 (97.4)	160 (98.8)				
The coach encouraged problem solving			**—**	**0.19**	**.55**	**.028**
Disagree/neutral	17 (2.3)	2 (1.2)				
Agree	725 (97.7)	160 (98.8)				
I could form a successful working relationship with the coach			**—**	**0.04**	**.07**	**.067**

Satisfaction-related factors[a]	Implementation, n (%)	Randomized controlled trial intervention, n (%)	Chi-square test (df)	Fisher exact test	*P* value	Effect size[b]
Disagree/neutral	24 (1.2)	1 (0.6)				
Agree	719 (96.8)	161 (99.4)				

[a] Disagree/neutral combines strongly disagree, disagree, and not agree or disagree. Agree combines agree and strongly agree.

[b] The effect sizes are measured by Cramér V.

[c] Not applicable.

Discussion

Principal Findings

This study describes the content and process of implementing a Web-based parent training program with telephone coaching in comparison with the RCT intervention. The program is unique because it is based on screening children from the general population during routine child health clinic checkups at the age of 4 years. First, the findings show that the characteristics of families recruited to the RCT study and the implementation groups were very similar. Second, the duration and content of the Web-based training and phone coaching were similar in both groups. Third, the satisfaction rates did not differ between the groups. Finally, against expectations, the discontinuation rate of the program was higher in the RCT group than in the implementation group.

The screening procedures that were used to identify children at risk during both the RCT and implementation phases resulted in similar profiles for families and child psychopathology. The only exceptions to the profiles were that the implementation group had 2-fold more moderate or severe difficulties (53,1%, 467/880) and more severe impairment based on SDQ impact score than the RCT intervention group (44.4%, 103/232). This result may indicate that the parents of the children with more severe disruptive behavior are more motivated to seek help and to keep up with the program than those of the children with minor behavioral problems. The same outcome has been found previously [24]. We suggest that the parental motivation related to the volume of the problem may act as one of the implementation drivers.

The study documented the duration and content of the 2 key elements of the remote parent training program, which were Web-based training and phone coaching. Both sets of parents spent about 80 min engaged in the program each week, when the figures for website use and coaching were combined. However, this did not include the time parents spent practicing the positive training skills with the child during the program, which was the key goal of the Web-based content and phone coaching.

The study showed high and similar parent satisfaction rates in both groups. The level of satisfaction with the program, how the program affected parenting skills, and how the parents worked with the phone coaches remained very high when the program made the transition from the RCT to the implementation phase. Most of the parents felt they had been able to form a successful working relationship with the coach. Finally, the study shows that less than 12,7 (112/882) of parents discontinued the program before the tenth week in the implementation group, which was almost 2-fold lower (24,1%, 56/232) than that in the RCT study. This indicates that the Web-based parent training programs supplemented with telephone coaching may result in high success. In our study, we identified core implementation drivers to facilitate the implementation process. The intervention fidelity was ensured in the implementation phase, by centralization of the conduction and coaching of the intervention, systematic quality assurance enabled via a digital platform, and supervision. Moreover, both ongoing training of the primary health care staff and stakeholders and facilitative administration, that is, regular meetings, user-friendly reports, and involving media, were likely to lead to successful implementation.

To our knowledge, the SFSW intervention is the first Web-based parent training program to use population-based screening to select its subjects and provide telephone coaching. Other digital interventions have reported using automated messages or emails to interact with parents [8,12]. During the program, the coach and the parent formed a working relationship, which was crucial for meeting mutually agreed goals [32,33]. Telephone coaching has a number of benefits: it enables real-time problem solving with the parent, provides direct feedback on how the parents adopt strategies and skills, and motivates the parent to continue the learning process. In addition, the website tracking and telephone coaching worked together to ensure sufficient exercises were conducted by the parent between the weekly sessions.

The findings of this study are important in a number of ways. They can help to inform best practice in implementing interventions that target large numbers of parents in a wide geographical area. They can also help to tackle the problems that are inherent with traditional parent training group interventions, such as drop-out rates and practical and resource issues. We have demonstrated that it is possible to replicate the success of an RCT at the implementation stage if the families are motivated and the structure of the program provides the right elements and meets their needs. Our experience also underlines that the ongoing supervision of those delivering the coaching element of the training is essential. It is notable that the use and adherence rates of digital mental health interventions might be lower in real-world settings, as implemented outside of research settings [34]. Our findings show that discontinuation of the intervention in the implementation group was exceptionally low and users were highly satisfied with the intervention.

Web-based interventions provide various opportunities for extending services to people with mental health [9] and behavioral problems, as they provide an effective way of reaching people, including the parents of children, as with our program. Moving interventions outside traditional clinics and into people's daily environments, like the internet, can facilitate better access to mental health services [14]. Web-based interventions can also remove the barriers associated with the face-to-face interventions and enable people to seek help for mental health problems without fear of being stigmatized [7,35,36]. From the families' point of view, digitally delivered programs can offer faster and more flexible services without the need for transport, juggling work schedules, arranging childcare, or the practical cost of accessing services. On the basis of the evidence, Web-based telephone-assisted interventions for children with disruptive behavioral problems could play a significant role in parent training by engaging families into low-threshold primary care services to promote the mental well-being of families and children [10,14,15].

As nearly all Finnish families use the child health clinic services, our treatment model reaches a very high proportion of the population. This enables early access to the low-threshold intervention, which is in accordance with the core principles of the Finnish child health clinic system, that is, early prevention, early detection, and increasing the awareness of physical and mental health. The elements of the treatment model constitute a new form of preventive intervention for disruptive behavior.

Limitations

There were some limitations to our study. We did not have any information about changes in the children's psychiatric problems or changes in parenting skills in the implementation group, which would have told us if the real-life implementation was as effective as the RCT study setting. We only had information about the time they spent on the website and telephone coaching, their satisfaction with the program, and how well they felt it met their needs. It must be noted that neither the therapeutic alliance nor the parent satisfaction questionnaire was validated for the Finnish settings. The alliance questionnaire was based on the original WAI (Working Alliance Inventory), whereas the satisfaction questionnaire was self-composed. Both the Web-based element and the coaching calls encouraged the parents to practice the skills they learnt in real life, but we do not have information about how much the parents actually did this. Information from public health nurses would have helped us to evaluate how the intervention was integrated into the primary health context. Finally, it is important to note that this study is an implementation study and not a sustainability study. However, the implementation sample in this study has been gathered up to 3 years after the RCT study ended. Sustaining an intervention in practice over time is different from scaling up interventions in practice. To maintain the effects observed in this study over a longer period may require additional strategies.

Conclusions

This study has important implications for planning low-cost, low-threshold, early, population-based, and evidence-based interventions in future. Our study revealed very similar findings between the RCT and implementation groups with regard to the profiles of the families and child psychopathology, the duration and content of the Web-based training and phone coaching, and the parents' satisfaction rates. The discontinuation rate was lower and disruptive behavior was more severe in the implementation group. These findings may suggest that the parents of the children with more severe disruptive behavior are highly motivated to receive help and complete the program. In this sense, parental motivation may be seen as an implementation driver. To our understanding, the successful implementation in this study constituted of numerous factors and their seamless interaction: the convenient and easy-to-use Web-based format with supplementary telephone coaching by trained professionals, sufficient training for stakeholders and primary health staff on the procedures, the psychosocial evaluation, and the support that the program families would receive. Moreover, we suppose that the central role is played by the primary health care, which in Finnish child health clinics has a potential to reach almost 100% of the children. We believe that this program provides a number of advantages over traditional parent training models, which face enormous challenges such as practical, resource, and retention issues, and that it could provide a blueprint for other cost-effective preventive and early intervention child mental health programs.

Acknowledgments

The research was funded by the Strategic Research Council of the Academy of Finland (Decision No. 303581), the Academy of Finland (Decision No. 320162), the Canadian Institutes of Health Research, the Jane and Aatos Erkko Foundation, the Association of Friends of the University Children's Hospitals, the Turku University Hospital Expert Responsibility Area Research Funding, and a Canada Research Chair.

Conflicts of Interest

None declared.

References

1. Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, et al. The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children. Health Technol Assess 2005 Dec;9(50):iii, ix-x, 1-233 [FREE Full text] [Medline: 16336845]

2. Dretzke J, Davenport C, Frew E, Barlow J, Stewart-Brown S, Bayliss S, et al. The clinical effectiveness of different parenting programmes for children with conduct problems: a systematic review of randomised controlled trials. Child Adolesc Psychiatry Ment Health 2009 Mar 04;3(1):7 [FREE Full text] [doi: 10.1186/1753-2000-3-7] [Medline: 19261188]
3. Eyberg SM, Nelson MM, Boggs SR. Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. J Clin Child Adolesc Psychol 2008 Jan;37(1):215-237. [doi: 10.1080/15374410701820117] [Medline: 18444059]
4. Perrin EC, Leslie LK, Boat T. Parenting as primary prevention. JAMA Pediatr 2016 Dec 01;170(7):637-638. [doi: 10.1001/jamapediatrics.2016.0225] [Medline: 27182902]
5. Jensen PS, Goldman E, Offord D, Costello EJ, Friedman R, Huff B, et al. Overlooked and underserved: "action signs" for identifying children with unmet mental health needs. Pediatrics 2011 Nov;128(5):970-979 [FREE Full text] [doi: 10.1542/peds.2009-0367] [Medline: 22025589]
6. 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]
7. Kazdin AE, Rabbitt SM. Novel models for delivering mental health services and reducing the burdens of mental illness. Clin Psychol Sci 2013 Jan 23;1(2):170-191. [doi: 10.1177/2167702612463566]
8. Breitenstein SM, Gross D, Christophersen R. Digital delivery methods of parenting training interventions: a systematic review. Worldviews Evid Based Nurs 2014 Jun;11(3):168-176. [doi: 10.1111/wvn.12040] [Medline: 24842341]
9. Barak A, Grohol JM. Current and future trends in internet-supported mental health interventions. J Technol Hum Serv 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
10. Taylor TK, Webster-Stratton C, Feil EG, Broadbent B, Widdop CS, Severson HH. Computer-based intervention with coaching: an example using the Incredible Years program. Cogn Behav Ther 2008;37(4):233-246 [FREE Full text] [doi: 10.1080/16506070802364511] [Medline: 18803072]
11. Hammond GC, Croudace TJ, Radhakrishnan M, Lafortune L, Watson A, McMillan-Shields F, et al. Comparative effectiveness of cognitive therapies delivered face-to-face or over the telephone: an observational study using propensity methods. PLoS One 2012;7(9):e42916 [FREE Full text] [doi: 10.1371/journal.pone.0042916] [Medline: 23028436]
12. Webster-Stratton C, Kolpacoff M, Hollinsworth T. Self-administered videotape therapy for families with conduct-problem children: comparison with two cost-effective treatments and a control group. J Consult Clin Psychol 1988 Aug;56(4):558-566. [Medline: 3143750]
13. Official Statistics of Finland (OSF). Statistics Finland. Helsinki: Statistics Finland; 2017. Use of information and communications technology by individuals URL: https://www.stat.fi/til/sutivi/index_en.html [accessed 2018-02-07] [WebCite Cache ID 6x3O76jFw]
14. 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]
15. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: a randomized clinical trial. JAMA Psychiatry 2016 Apr;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]
16. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Hinkka-Yli-Salomäki S, et al. Two-year follow-up of internet and telephone assisted parent training for disruptive behavior at age 4. J Am Acad Child Adolesc Psychiatry 2018 Sep;57(9):658-668.e1. [doi: 10.1016/j.jaac.2018.07.001] [Medline: 30196869]
17. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Dec 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
18. Gyllenberg D, Gissler M, Malm H, Artama M, Hinkka-Yli-Salomäki S, Brown AS, et al. Specialized service use for psychiatric and neurodevelopmental disorders by age 14 in Finland. Psychiatr Serv 2014 Mar 01;65(3):367-373 [FREE Full text] [doi: 10.1176/appi.ps.201200544] [Medline: 24337256]
19. World Health Organization. 2018. Adolescents: health risks and solutions URL: http://www.who.int/en/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2018-06-21] [WebCite Cache ID 70KqWlNAj]
20. Michelson D, Davenport C, Dretzke J, Barlow J, Day C. Do evidence-based interventions work when tested in the. Clin Child Fam Psychol Rev 2013 Mar;16(1):18-34. [doi: 10.1007/s10567-013-0128-0] [Medline: 23420407]
21. Fixsen DL, Blase KA, Naoom SF, Wallace F. Core implementation components. Res Soc Work Pract 2009 May 27;19(5):531-540. [doi: 10.1177/1049731509335549]
22. Dobson KS. Definitional and practical issues in the assessment of treatment integrity. Clin Psychol: Sci Pr 2005 Dec 01;12(4):384-387. [doi: 10.1093/clipsy/bpi046]
23. Sánchez V, Steckler A, Nitirat P, Hallfors D, Cho H, Brodish P. Fidelity of implementation in a treatment effectiveness trial of reconnecting youth. Health Educ Res 2007 Feb;22(1):95-107. [doi: 10.1093/her/cyl052] [Medline: 16807378]
24. Fossum S, Ristkari T, Cunningham C, McGrath PJ, Suominen A, Huttunen J, et al. Parental and child factors associated with participation in a randomised control trial of an internet-assisted parent training programme. Child Adolesc Ment Health 2016 Nov 10;23(2):71-77. [doi: 10.1111/camh.12193]

25. National Institute for Health and Welfare. Department of Health and Welfare. Child welfare clinic URL: https://www.thl.fi/en/web/lapset-nuoret-ja-perheet/peruspalvelut/aitiys_ja_lastenneuvola/lastenneuvola [accessed 2018-02-07] [WebCite Cache ID 6x3Pvhh8p]
26. McGrath PJ, Sourander A, Lingley-Pottie P, Ristkari T, Cunningham C, Huttunen J, et al. Remote population-based intervention for disruptive behavior at age four: study protocol for a randomized trial of Internet-assisted parent training (Strongest Families Finland-Canada). BMC Public Health 2013 Oct 21;13:985 [FREE Full text] [doi: 10.1186/1471-2458-13-985] [Medline: 24139323]
27. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
28. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, StaRI Group. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. BMJ Open 2017 Dec 03;7(4):e013318 [FREE Full text] [doi: 10.1136/bmjopen-2016-013318] [Medline: 28373250]
29. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-586. [Medline: 9255702]
30. Koskelainen M, Sourander A, Kaljonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. Eur Child Adolesc Psychiatry 2000 Dec;9(4):277-284. [Medline: 11202103]
31. Borg A, Kaukonen P, Salmelin R, Joukamaa M, Tamminen T. Reliability of the strengths and difficulties questionnaire among Finnish 4-9-year-old children. Nord J Psychiatry 2012 Dec;66(6):403-413. [doi: 10.3109/08039488.2012.660706] [Medline: 22397524]
32. Lingely-Pottie P, McGrath PJ. A therapeutic alliance can exist without face-to-face contact. J Telemed Telecare 2006;12(8):396-399. [doi: 10.1258/135763306779378690] [Medline: 17227604]
33. Lingley-Pottie P, McGrath PJ. Distance therapeutic alliance: the participant's experience. ANS Adv Nurs Sci 2007 Dec;30(4):353-366. [doi: 10.1097/01.ANS.0000300184.94595.25] [Medline: 18025870]
34. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
35. Bee PE, Bower P, Lovell K, Gilbody S, Richards D, Gask L, et al. Psychotherapy mediated by remote communication technologies: a meta-analytic review. BMC Psychiatry 2008 Jul 22;8:60 [FREE Full text] [doi: 10.1186/1471-244X-8-60] [Medline: 18647396]
36. Slone NC, Reese RJ, McClellan MJ. Telepsychology outcome research with children and adolescents: a review of the literature. Psychol Serv 2012 Aug;9(3):272-292. [doi: 10.1037/a0027607] [Medline: 22867120]

Abbreviations

OR: odds ratio
RCT: randomized controlled trial
SDQ: Strengths and Difficulties Questionnaire
SFSW: Strongest Families Smart Website

Edited by G Eysenbach; submitted 29.06.18; peer-reviewed by F Drozd, J Doty; comments to author 16.08.18; revised version received 09.11.18; accepted 31.12.18; published 11.04.19.

Please cite as:
Ristkari T, Kurki M, Suominen A, Gilbert S, Sinokki A, Kinnunen M, Huttunen J, McGrath P, Sourander A
Web-Based Parent Training Intervention With Telephone Coaching for Disruptive Behavior in 4-Year-Old Children in Real-World Practice: Implementation Study
J Med Internet Res 2019;21(4):e11446
URL: https://www.jmir.org/2019/4/e11446/
doi:10.2196/11446
PMID:30973337

Original Paper

Integration of Rehabilitation Activities Into Everyday Life Through Telerehabilitation: Qualitative Study of Cardiac Patients and Their Partners

Birthe Dinesen[1*], MSc, PhD; Gitte Nielsen[2*], MD, PhD; Jan Jesper Andreasen[3,4*], MD, PhD; Helle Spindler[5*], MSc, PhD

[1]Laboratory of Welfare Technologies - Telehealth and Telerehabilitation, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[2]Department of Cardiology, Region Hospital North Jutland, Hjoerring, Denmark

[3]Department of Cardiothoracic Surgery, Aalborg University Hospital, Aalborg, Denmark

[4]Department of Clinical Medicine, Aalborg University Hospital, Aalborg, Denmark

[5]Department of Psychology and Behavioral Sciences, Aarhus University, Aarhus, Denmark

[*]all authors contributed equally

Corresponding Author:
Birthe Dinesen, MSc, PhD
Laboratory of Welfare Technologies - Telehealth and Telerehabilitation
Department of Health Science and Technology
Aalborg University
Frederik Bajers vej 7
Aalborg, 9220
Denmark
Phone: 45 20515944
Email: bid@hst.aau.dk

Abstract

Background: Implementation of cardiac rehabilitation has not been optimal, with patient participation rates below 50%. Factors that contribute to cardiac patients' lack of participation in rehabilitation programs are patient motivation, logistical difficulties in getting to the rehabilitation facilities, lack of psychosocial elements, and individualization of activities in the rehabilitation programs. Telerehabilitation has been proposed as a new way to address the challenge of engaging and motivating cardiac patients and their partners to participate in rehabilitation.

Objective: The aim of this study was to explore the experiences of cardiac patients and their partners of participating in the Teledialog Telerehabilitation Program (TTP). The Teledialog program consisted of a digital rehabilitation plan, transmission of health data from patient's home to hospital and health care center, and an interactive Web portal with information and training videos.

Methods: This case study used a theoretical approach combining the "community of practice" approach and self-determination theory. A triangulation of data collection techniques was used, including documents, participant observation (72 hours), and qualitative interviews with cardiac patients and their partners enrolled in the telerehabilitation group. A total of 14 cardiac patients, 12 patient spouses/partners, and 1 son participated in the study. The participants were interviewed at enrollment in the telerehabilitation program and after 12 weeks of participation in the program. Interview data were analyzed using NVivo 11.0.

Results: Patients and their partners found the Web portal ActiveHeart.dk and the electronic rehabilitation (e-rehabilitation) plan to be helpful tools for health education, coordinating rehabilitation goals, creating an overview of the data, and ensuring continuity in the rehabilitation process. The patients felt that the TTP treated them as individuals, gave them a sense of autonomy, and provided enhanced relatedness to health care professionals and partners and a sense of competence as active participants in their own rehabilitation process. Some patients missed being part of a community of practice with other cardiac patients and did not use the Web forum. Patients' partners found that the telerehabilitation program gave them a sense of security and helped them balance their involvement as a partner to the patient and not push the patient too hard.

Conclusions: Cardiac patients and their partners found telerehabilitation technologies a useful digital toolbox in the rehabilitation process. Telerehabilitation motivated the patients to integrate rehabilitation activities into their work schedule and everyday life and made them feel like unique individuals. Participating in the Teledialog Telerehabilitation Program might not be a suitable strategy for all cardiac patients. Being a patient's partner in the telerehabilitation program was associated with a heightened sense of security, navigation between active involvement in the rehabilitation process, being an equal partner, and not pushing the patient too hard.

(J Med Internet Res 2019;21(4):e13281) doi:10.2196/13281

KEYWORDS

telerehabilitation; heart diseases; social media; qualitative study

Introduction

Each year, 17.9 million people die from cardiovascular diseases, and it is estimated that cardiovascular diseases cause 31% of all deaths worldwide [1]. Cardiac rehabilitation programs include interventions such as exercise and patient education on risk factors [2-3]. Cardiac rehabilitation helps patients alter their lifestyle and decrease their mortality rate [2-4]. However, despite international recommendations, effective implementation of cardiac rehabilitation after cardiovascular diseases has not been optimal, with participation rates below 50% [1-3]. Many factors contribute to patients' lack of participation in rehabilitation activities including patients' motivation, lack of means of transport to the clinic, time constraints, scheduling commitments associated with returning to work, lack of psychosocial counseling in the programs, lack of individualization of activities in the rehabilitation programs, and lack of involvement of partners [3,5-7]. More innovative models for cardiac rehabilitation are needed to address the challenge of engaging and motivating cardiac patients and their partners to participate in rehabilitation. Telerehabilitation can be one such new approach. Telerehabilitation is defined as the delivery of rehabilitation services via information and communication technologies [8].

A review of internet-based cardiovascular rehabilitation from 2013 [9,10], including a small number of trials and few outcome measures (physical activity, clinical, and psychosocial outcomes), reported positive results with regard to patient outcome and feedback. However, none of the studies were integrated with clinical practice, as confirmed by another review from 2013 [11].

A 2015 review [7] on telerehabilitation among cardiac patients found that several more studies had been carried out, but these studies were characterized as heterogeneous with respect to interventions, patients, and outcome measures. Most interventions had only one or two components of cardiac rehabilitation, and physical activities were the most frequently used approach [7]. Systematic reviews on cardiac telerehabilitation show that it can help reduce depression and improve both functional capacity and physical activity level [12,13].

There is a lack of qualitative studies exploring cardiac rehabilitation participants' and their partners' experiences in telerehabilitation programs. Partners of cardiac patients play an important role in patients' adjustment to lifestyle changes and the rehabilitation process [14,15]. Qualitative knowledge about patients' and their partners' experiences of telerehabilitation can be a valuable source for optimizing future cardiac telerehabilitation programs.

From 2011 to 2012, the Danish Teledialog Telerehabilitation Program (TTP) for cardiac patients and their partners was developed through user-driven innovation [16]. The TTP emphasized collaboration between cardiac patients, their partners/relatives, health care professionals, researchers, and representatives from companies. The TTP was developed with inspiration from learning theory, specifically the "communities of practices" approach [17]. The goal was to facilitate mutual learning between patients, partners, and health care professionals in the telerehabilitation program. Self-determination theory (SDT) [18,19] was also applied, with the purpose of exploring whether telerehabilitation could help engage and motivate patients and partners in the rehabilitation process.

The aim of this study was to explore the experiences of cardiac patients and their partners who participated in the TTP.

Methods

Design

This study is a substudy of the TTP and was performed as a descriptive case study [20]. According to Yin (2013), a case study is defined as "an empirical inquiry that investigates a contemporary phenomenon (the 'case') in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident" [20].

Description of Sampling

The inclusion criteria for enrolling cardiac patients in the Teledialog project were age over 18 years; living in North Jutland in Denmark; history of acute coronary syndrome, heart failure, or coronary artery bypass surgery/valve surgery; living in an area with mobile coverage; and user-level competence in information and communication technologies. The exclusion criteria were inability to speak Danish, pregnancy, or breastfeeding. From May 2013 to January 2014, patients enrolled in the telerehabilitation group of the Teledialog trial were consecutively contacted and invited for interviews. A total of 14 cardiac patients (9 men and 5 women) participated in two interviews: One, at enrollment into the TTP and another, after 12 weeks of participation in the program. A total of 12 patient partners and one son of a patient participated in the interviews. Table 1 shows the baseline characteristics of the interviewed patients and partners.

Table 1. Basic characteristics of the interviewed patients and partners from the telerehabilitation group.

Variables	Interviewed patients in the telerehabilitation group (n=14)	Interviewed partners (n=13)
Age (years), mean (SD); range		
Men[a]	61.56 (11.39); 47-85	54 (5.06); 45-60
Women[b]	58.40 (10.62); 51-77	58,25 (12.27); 40-82
Total	60.43 (10.82); 47-85	56,62 (10.33); 40-82
Weight (kg)	90.03 (23.706); 43.90-130.00	N/A[c]
Blood pressure, mean (SD); range		N/A
Systolic blood pressure	125.93 (17.76); 90-95	N/A
Diastolic blood pressure	73.71 (11.22), 58-90	N/A
Heart rate (beats/minute), mean (SD), range	70.143 (7.02), 57-81	N/A
Condition, n (%)		
Acute coronary syndrome	7 (50)	N/A
Cardiac surgery	4 (28.57)	N/A
Heart failure	3 (21.43)	N/A
Arterial sclerosis and heart failure	0 (0)	N/A
Status, n (%)		
Single	1 (7.14)	1[d] (7.69)
Married or living with a partner	13 (92.85)	12 (92.30)
Education, n (%)		
Elementary school	2 (14.28)	3 (23.07)
High school	3 (21.42)	2 (15.38)
Skilled work	8 (57.14)	6 (46.15)
Higher education	0 (0)	1 (7.69)
Bachelor’s degree	0 (0)	0 (0)
Master’s degree	1 (7.14)	1 (7.69)
Employment, n (%)		
<20 h/week	0 (0)	0 (0)
20-36 h/week	0 (0)	0 (0)
37 h/week	4 (28.54)	7 (53.84)
On sick leave	5 (35.71)	0 (0)
Unemployed	0 (0.00)	1(7.69)
Retired	5 (35.71)	5 (38.46)
Missing data	0 (0)	0 (0)
Participating institutions, n (%)		
Hospital	6 (42.85)	N/A
Health care center	7 (50.00)	N/A
Call center	1 (7.14)	N/A

[a] n=9 (patients); n=5 (partners).

[b] n=5 (patients); n=8 (partners).

[c] N/A: not applicable.

[d] Son was interviewed.

Prior to the initial interview, the patients were asked to invite a partner to participate in the interviews. One patient who was single invited a son to participate, as the son was an important part of the patient’s life after cardiac disease.

None of the patients refused to participate in the two interviews or dropped out during the intervention period.

Three patients within the total intervention group (n=75) dropped out during the 12-week telerehabilitation program due to disconfirmation of diagnosis (n=1), inability to cope with the project (n=1), and serious illness (n=1).

Presentation of the Teledialog Telerehabilitation Program

The TTP followed the current guidelines for cardiac rehabilitation developed by the Danish Health Agency [21,22] and the European Association of Cardiovascular Prevention and Rehabilitation [23]. The overall aims of the TTP were to create more personalized rehabilitation; engage and motivate more cardiac patients to participate in rehabilitation; increase patients' quality of life; and facilitate coherence between patients, partners, and health care professionals in the rehabilitation process.

Textbox 1 provides an overview of the content (equipment, exercise, and education) of the TTP, which lasted 12 weeks for each patient.

In the TTP, the participating institutions were a cardiology ward at a regional hospital, a thoracic ward at a university hospital, a call center, and four health care centers in two municipalities that participated.

At discharge from the hospital, the patients and their partners were invited to a meeting with a project nurse. At the meeting, the patient was interviewed to identify their specific needs for rehabilitation. An individualized rehabilitation plan was designed in collaboration with the patients and their partners. Patients could participate in telerehabilitation from hospital, health care center, or call center. They were also allowed to participate in rehabilitation activities at the health care center, for example, as group-based training. They received education on using the devices for measuring blood pressure, pulse, weight, and steps and were taught to navigate the ActiveHeart Web portal, which was the center of the telerehabilitation technologies (Textbox 1, Figure 1). The ActiveHeart is an interactive information site with text and videos on rehabilitation issues. The portal also has a Web forum that enables patients and their partners to communicate with each other. The function of the e-rehabilitation plan is described in Textbox 1. Patients, partners, nurses, doctors from the hospital, physiotherapists, and nurses from the health care center had access to the e-rehabilitation plan. The patient had to give written consent to allow the parties access to the e-rehabilitation plan. During the 12 weeks of rehabilitation, the parties used the e-rehabilitation plan for communication, sharing data on home measurements, information, and goal setting for rehabilitation activities. Each patient had a contact person who was responsible for the collaboration with the individual patient and his/her rehabilitation process.

A video of the TTP is provided in Multimedia Appendix 1. The TTP was tested from 2012 to 2014 in a randomized controlled trial.

Textbox 1. The Teledialog Telerehabilitation Program content.

Equipment:

- Telehealth monitor approved for medical use and paired with a telehealth monitor in advance
- Sphygmomanometer approved for medical use and paired with a telehealth monitor in advance
- Weight scale approved for medical use and paired with a telehealth monitor in advance
- Device for measuring electrocardiography approved for medical use and paired with a telehealth monitor in advance
- Digital pedometer FitBit Ultra registered the number of steps
- ActiveHeart is an interactive Web portal with information on heart functions, heart diseases and symptoms, videos with instructions on exercises, and brief rehabilitation narratives by patients and relatives. There was also a Web forum enabling patients to communicate with each other. The Web forum was moderated by a nurse. ActiveHeart.dk was developed by Aalborg University.
- Electronic rehabilitation plan (digital rehabilitation plan) giving an overview of patient data including goal and plan for rehabilitation, appointments, diary, and an overview of measured values (pulse, blood pressure, weight, and steps). Patients, relatives, and health care professionals had access to the plan. The Shared Care Platform was developed by IBM Corp.
- Android tablet (Samsung Galaxy Tab 2, 10.1) for accessing the interactive Web portal and digital rehabilitation plan.

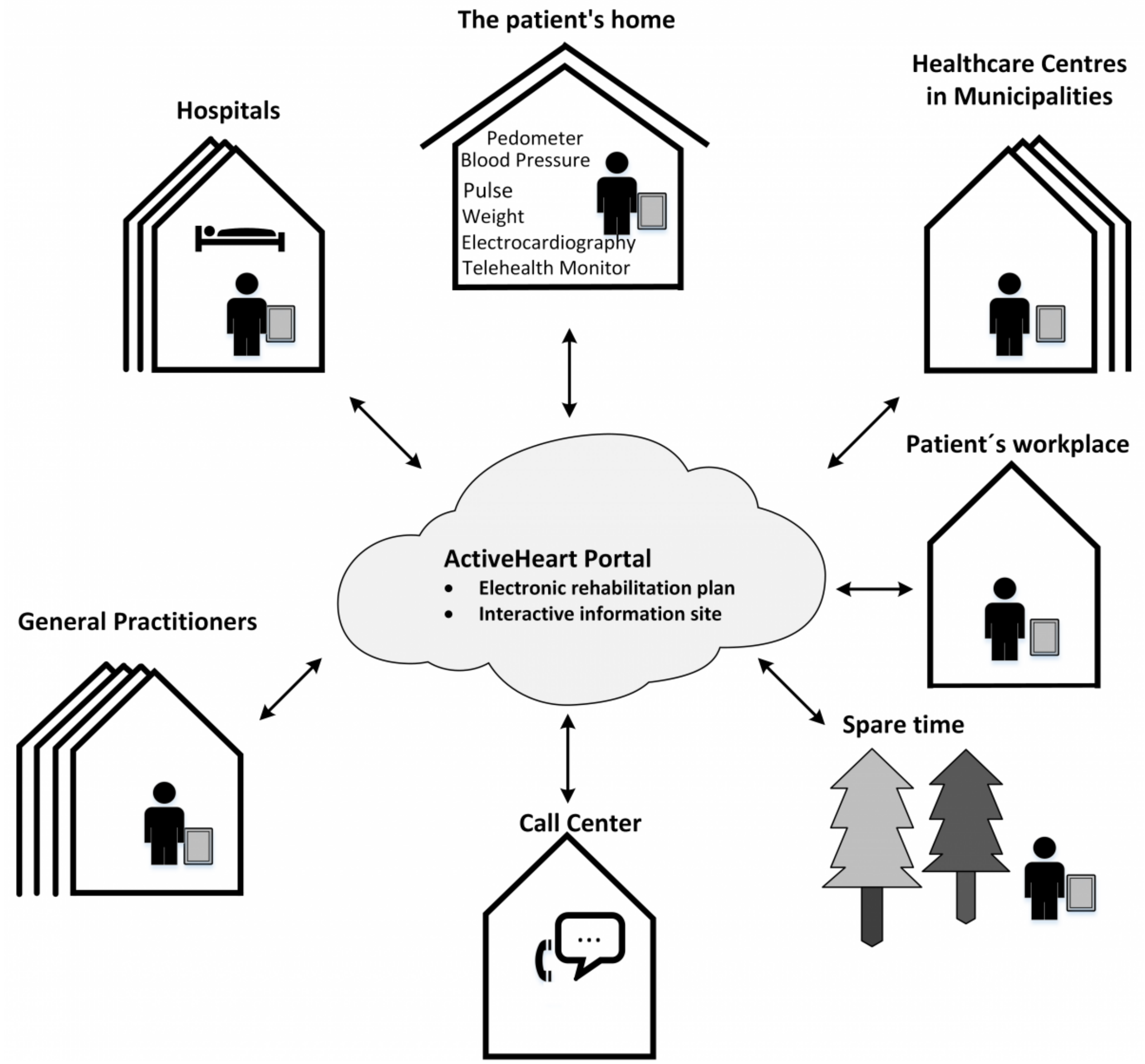
Exercise:

- Individual and group-based aerobic and strength training

Education:

- Individual and group-based education within the following themes: self-management, physical activity, nutritional counseling, medications, psychosocial support, and managing a new lifestyle

Figure 1. Overview of the Teledialog Telerehabilitation Context.



Theory

The theoretical framework for this case study consists of the "community of practice" approach [17] and SDT [18,19]. Wenger defined a "community of practice" as a group of people who share a concern or passion for something they do and who interact on a regular basis. This approach was applied to this study to determine whether the TTP can stimulate the emergence of communities of practice among patients, partners, and health care professionals.

SDT [24] focuses on human motivation, in general, and more specifically, whether a telerehabilitation program based on SDT can help motivate cardiac patients and their partners to improve their disease management skills and enhance their participation in rehabilitation. SDT describes how human beings inherently strive to develop or grow psychologically and how external factors may either impede or support this process. As such, SDT also provides a theoretical framework for understanding long-term motivation, as extrinsic motivation (ie, being prompted by others to perform a certain behavior) persists only as long as the external motivator is present. In contrast, intrinsic motivation refers to internalized motivation (ie, no external motivators are needed for the motivation to persist). To experience intrinsic motivation, it is necessary that all three basic human needs are sufficiently fulfilled: autonomy (acting in accordance with one's internal values), competency (having the necessary knowledge and skills), and relatedness (being recognized by others) [18,19]. Motivation can be nurtured by supporting the patient's experience of having these three basic needs sufficiently fulfilled simultaneously.

Key themes from the community of practice and SDT theories were applied in the observation and interview guides used in the data collection process (Multimedia Appendix 2, Multimedia Appendix 3).

Data Collection Techniques

A triangulation of data collection techniques was used, such as documentary materials, participant observation, and qualitative interviews with patients and partners. Multiple sources of evidence were used to help provide multiple measures of the same phenomenon [20] in the case study. Data from documentary materials, participant observation, and interviews were analyzed in the same process, as the triangulation of data helped enhance the validity of the case study.

Document Analysis

As background for the study, documents and reports on rehabilitation strategies, policies, and homepages from the involved health care organizations were reviewed. The aim of using documents was to achieve a basic understanding of the context of cardiac rehabilitation activities for the case study. Documents were analyzed using NVivo 11.0 (QSR International, Melbourne, Australia).

Participant Observation

An observational guide (Multimedia Appendix 2) was developed based on the theoretical framework. The aim of performing participant observation [25] was to observe the interaction between patients and health care professionals in their telerehabilitation activities, patients' participation in the TTP, and patients' interaction with their partners using the technologies. Observations took place during meetings between patients and health care professionals on discharge from the hospital, when patients were introduced to the telerehabilitation program and technologies and at the commencement of rehabilitation carried out at the health care center. Observations also took place in patients' homes, focusing on the patients' use of the telerehabilitation technologies in everyday life and in their interaction with partners. The majority of the observations were carried out in the patients' homes.

The observations were conducted by the first author (BD), who has a nursing and social science background, and the last author (HS), who has a background in Psychology. A total of 72 hours of participant observation was carried out. Notes were taken and documented in a Microsoft Word (Redmond, WA) file and analyzed using NVivo 11.0.

Interviews with Patients and Partners

As per the study by Kvale and Brinkmann [26], semistructured qualitative interviews were conducted with cardiac patients and relatives, first at the time of enrollment into the TTP and then after 12 weeks of participation. The aim of the initial interviews was to gain an understanding of the patients' disease and rehabilitation plan and the everyday life of the patient and his/her partner. The aim of interviews after 12 weeks of participation in the TTP was to explore the experiences of the cardiac patients and their partners.

Both patients and their partners were interviewed at the same time to ensure openness in the situation, although we were aware that it could be a barrier for some patients and their partners. One (single) patient wanted a son to represent a "partner's point of view" due to their close daily relationship. In order to exclude topics that the patient and partners might not want to discuss together, the patient and partner were told prior to the interview that they could take a break in the interview to postpone talking about the topic until later when they could talk to the researcher on the phone, or to refuse talking about a topic. No patients or partners (nor the son) asked to delay talking about topics.

On enrollment, the main dimensions of the interviews were introduction; everyday life, limitations imposed by the disease on everyday life, family life, experience with use of technology, expectation to use of telerehabilitation technologies, social network, management of illness in everyday life, co-operation with partner, co-operation with health care professionals, and expectation for the future. After the patients had participated in the TTP for 12 weeks, the main dimensions in the interview guide for both the patient and their partner were influence of TTP on daily life, possibilities and limitation taking measurements, experiences with use of technologies (e-rehabilitation plan, step-counter, and ActiveHeart), social network (including other cardiac patients), managing one's own rehabilitation supported by technology (information about rehabilitation, goal setting, and co-operation with partner and health care professionals), and experiences in participating in the TTP. A sample interview guide is provided in Multimedia Appendix 3.

The interviews were conducted by the first (BD) and last author (HS). The researchers had no relationship with the patients or their partners prior to commencement of the study. The interviews were tape recorded and transcribed by a research assistant. The interviews took place in the homes of the patients and lasted between 90 and 120 minutes. Interviews were conducted until we reached the point of data saturation.

Analysis

All data from documents, notes from participant observations, and interviews were analyzed using NVivo 11.0. The first (BD) and last authors (HS) designed the code tree and performed the analysis. As per the study by Kvale and Brinkman [26], the analysis was carried out in the following steps:

1. Key themes and definitions from the theoretical framework were identified by the authors.
2. All data from documents, observations, and interviews were reviewed to obtain an overall impression of the themes on how the cardiac patients and their partners experienced participation in the TTP.
3. Based on key definitions and concepts derived from the theoretical framework and from reviewing observations and interviews, a code tree was designed (Textbox 2).
4. Before coding the data, the code tree was reviewed and discussed by the first (BD) and last authors (HS) to ensure intersubjectivity.
5. In the first step of the coding process, it was important to gain an understanding of the patient's and his/her partner's self-understanding.
6. In the next step, the interpretation was widened to include an understanding beyond the patient and their partner. What motivated their views and what was the scope of action?
7. In the final phase, the data were analyzed with a focus on the experiences of cardiac patients and their partners in participating in the TTP.
8. The data have been condensed and are presented in Textboxes 3 and 4 in the form of key themes and findings.

Textbox 2. Code tree.

Use of technology in everyday life:

- Use of ActiveHeart Web portal
- Electronic rehabilitation plan

Participating in the Teledialog Telerehabilitation Program as a patient (autonomy):

- Feeling of self-control

Feeling of relatedness:

- Partner
- Family
- Health care professionals
- Other cardiac patients

Competencies:

- Technologies as a toolbox
- Empowerment in everyday life

Partner’s view:

- Use of ActiveHeart
- Rehabilitation activities and e-rehabilitation plan
- Being a partner

Textbox 3. Findings from cardiac patients participating in the Teledialog project.

Use of Technology:

- ActiveHeart.dk was used as a “virtual book” on education.
- Electronic rehabilitation plan provided active assistance in defining and continuously reviewing goals and plans for rehabilitation activities and ensuring continuity in the rehabilitation process.

Autonomy:

- Flexibility and possibilities for individual decision making in the Teledialog Telerehabilitation Program.
- The Teledialog Telerehabilitation Program encouraged patients to carry out activities with their own initiative.

Relatedness:

- Health care professionals become coaches in aiding patients’ return to everyday life.
- Partners were considered important sources of support for continuing training.
- No feeling of being part of a community of practice with other cardiac patients.
- Web forum was not used because it became too private.

Competences:

- Telerehabilitation technologies and data overview encouraged patients to integrate rehabilitation into their everyday life.
- Patients were better able to perform rehabilitation activities outside hospitals and health care centers despite the need to plan time and geographical distance.

Textbox 4. Findings from the partners participating in the Teledialog project.

Use of technologies:

- ActiveHeart.dk is a useful tool for the patient’s partner.
- E-rehabilitation plan is helpful in coordinating goals and creating overview of partner’s rehabilitation activities.

Being a partner:

- Mutual agreement on goals for planned rehabilitation activities facilitated coherence
- Sharing online information created a sense of security

- Partners took too much responsibility reviewing data on behalf of their partner
- Balance between being overly involved and being an equal partner
- Pushing the patient too “fast forward” in the rehabilitation process

Ethical Approval

The project was approved by the Danish Ethical Committee (N-20120051) and performed according to the Declaration of Helsinki. All participants signed an informed consent agreement. The Teledialog Project is registered at ClinicalTrials.gov (NCT01752192). The study followed the guidelines in the Danish Act on Processing of Personal Data.

Results

Key Themes and Findings

In Textboxes 3 and 4, key themes and findings from patients and their partners are listed and described in more detail with quotations from patient and partner interviews.

We found no differences in perceptions based on sex, age, type of disease, or illness among the cardiac patients and their partners. As shown in Table 1 on baseline characteristics, the men and women among patients and their partners were within the same approximate age group.

Use of Technology

Both patients and their partners expressed the view that the ActiveHeart.dk was a useful tool for patient education in the rehabilitation process and that it provided them with relevant information in text and video about heart diseases, symptoms, and topics on changing lifestyle:

> *I have been able to find relevant information in the ActiveHeart.dk on topics about rehabilitation. I have used the videos of how to do exercises very much, and the video with other cardiac patients describing how they were feeling has been helpful for me and my wife.* [Patient #15, male]
>
> *The e-plan has been the backbone of my rehabilitation process between me, the healthcare professionals and my wife.* [Patient #19, male]

A partner expressed her views about the ActiveHeart.dk:

> *It has been very helpful to have all information online and available 24/7, and especially the videos with narratives from partners have been very informative and made me feel secure.* [Partner #16, female]

The patients’ and partners’ views about the e-rehabilitation plan generally indicated that the plan had given them an overview of goals, plans, and appointments, creating a sense of coherence in the rehabilitation process:

> *It is a relief to have the digital record and be able to follow the goals and plans made for my husband and which encouraged him to stick to the plans.* [Partner #18, female]

Integrating patient education, communication with health care professionals, and the e-rehabilitation plan using a single access point, namely, the ActiveHeart.dk Web portal, provided patients and partners with the necessary knowledge and overview of the rehabilitation process, resulting in a much-valued sense of coherence in the rehabilitation program.

Autonomy

Our data indicate that the TTP encouraged the patients’ autonomy because they felt more involved in making personal decisions related to their rehabilitation process.

The patients felt they had control over their own situation. One patient reported the following in the interview:

> *The Teledialog project has provided the framework for individualized rehabilitation at a distance and stimulated me to do activities that I felt were useful for me in my situation.* [Patient #21, male]
>
> *Being able to view my own data and be ‘the captain’ of my own rehabilitation is important to me.* [Patient #23, female]

Relatedness

During their participation in the TTP, cardiac patients felt that the health care professionals took on a new role as coaches in aiding their return to everyday life. The coaching revolved around the mutually accessed data (steps, weight, blood pressure, and pulse):

> *I have experienced the healthcare professionals in a new role as coaches, and I think this is because they can see how physically active, I am, and in this way, encourage me to keep me focused on my rehabilitation.* [Patient #22, male]

The partners became an important support resource in everyday life for the patients, motivating them to keep up their training lifestyle changes:

> *My wife is an excellent supporter for me when I do not feel like taking my daily walk or going to the healthcare center for training.* [Patient #27, male]
>
> *I appreciate it when my husband keeps pushing and motivating me to keep training and following my goals for my rehabilitation.* [Patient #29, female]

As such, the TTP enabled both health care professionals and partners to provide important feedback to the patient, who, in turn, experienced enhanced relatedness.

Some patients felt that by participating in the TTP, they were lacking a sense of belonging to a patient community of practice. They missed the contact with other cardiac patients, where they could exchange views and experiences about their lifestyle changes. Observational notes showed that the patients who missed having more contact with other cardiac patients were first-time cardiac patients:

> *I enjoy the idea of telerehabilitation, but I would have liked more possibilities to meet with other cardiac patients in order to exchange experiences after admission to the hospital.* [Patient #29, female]

This suggests that the TTP was not able to build an online interactive forum in which patients could experience patient-to-patient relatedness. Only a few patients used the Web forum:

> *I do not like to communicate on a web forum. I think that my rehabilitation and questions related to this are a private matter.* [Patient #2, male]

Competences

Patients felt that the daily data summary they received was a good source of information about their competency with regard to their required rehabilitation activities. The data summary encouraged patients to perform their rehabilitation activities more comprehensively and more frequently, at any time or place, and independent of geographical distance (at home, work, or a summer house):

> *I get motivated because I can see how I perform on my physical activity and then carry out the activities whenever it fits into my daily schedule and working life.* [Patient #27, female]
>
> *Based upon [my rehabilitation] data, I feel motivated to keep focus on progress towards my goals and plans for my rehabilitation.* [Patient #16, male]

Being a Partner

The partners viewed the e-rehabilitation plan as helpful in facilitating an understanding and coherence in the rehabilitation process for both themselves and the patient, thus creating a feeling of security:

> *Being involved in the process of goal-setting for the rehabilitation of my husband gave me an understanding of the importance of changing lifestyle and the challenges that can arise.* [Partner #17, female]

Some partners took too much responsibility during their participation in the TTP. They found it difficult to navigate between the desire to be involved, to be over-protective of their spouse, and to remain an equal partner in the rehabilitation process. The result could be a situation where they pushed the patient too hard in the rehabilitation process. Observations showed that this tendency occurred mostly in female partners:

> *I think that in the beginning of my husband's rehabilitation process, I was very eager to help understand the data and define activities; e.g., walking more steps or losing weight. I felt that my husband became passive and was not engaged...we had a talk about my role, and I realized that I had become too involved. I think there is a balance about being involved and being an equal partner...I have now realized that rehabilitation takes time.* [Partner #20, female]
>
> *I had a difficult time balancing between motivating my husband to get going on a new lifestyle and at the same time not taking over… it resulted in me pushing my partner too hard, and he became frustrated.* [Partner #22, female]

Discussion

Principal Results

A triangulation of data was used in the case study, and the aim was twofold: (1) to provide multiple measures of the same phenomenon and (2) to construct validity of the case study.

The aim of the case study was to explore the experiences of cardiac patients and their partners in participation in the TTP. Findings showed that they found the technology, such as the ActiveHeart.dk Web portal and the e-rehabilitation plan, to be helpful tools in educating, coordinating (goal-setting), and creating an overview of their rehabilitation data and ensuring continuity in the rehabilitation process. The patients felt that the TTP make them feel like individuals and gave them a sense of autonomy, an increased relatedness to health care professionals and partners, and a sense of competence as active players in their own rehabilitation process. Some patients missed being part of a community of practice with other cardiac patients, and these patients did not use the Web forum option on ActiveHeart. Patients' partners found that the telerehabilitation program gave them a sense of security and helped them navigate between involvement in the rehabilitation, being an equal partner as the patient, and not pushing the patient too hard.

Interpreting the Findings in the Context of the Research Literature

Patients and health care professionals often have competing priorities and interests in terms of the kind of information they feel is most important during treatment [27]. Our findings highlight the fact that both patients and their partners found ActiveHeart and the e-rehabilitation plan to be useful tools in the rehabilitation process. The health care professionals from the Teledialog project stated that the e-rehabilitation plan facilitated knowledge and information sharing between themselves, the patients, and their partners [28]. Furthermore, the patients became active collaborators in their rehabilitation process [28]. Jansson et al [29] reviewed the use of health plans formulated between health care professionals and patients with acute coronary syndrome. Similar to our findings, they reported that shared information and shared decision making facilitate partnership between patients and health care professionals. In addition, we identified one ongoing study on cardiac telerehabilitation that used a personalized patient-centered Web app [30]. In this study, however, it was unclear whether the patients and their cohabiting partners could share information with health care professionals across sectors in the rehabilitation process, as was done in the Teledialog project.

In the Teledialog project, the patients did not seem to have any technological challenges, perhaps due to the inclusion criteria in this study, where the selected patients had to have user-level competence in information and communication technologies. In a substudy, the patients expressed the view that the ActiveHeart Web portal was easy to access, user friendly, and written in an understandable language [31].

Factors that might be obstacles to adoption of telehealth or telerehabilitation can be requirements for technical competences, operation of equipment, threats to identity, and attitudes that digital services undermine in self-care and independence [32]. A recent study on cardiac rehabilitees' experiences with technology concluded that they had different experiences and expectations and they expect the technology to be simple, flexible, and easy to use and learn [33]. Albert et al [34] highlighted the importance of matching technologies with patients' personal learning and use needs [34].

In our study, the cardiac patients felt that the TTP supported their sense of autonomy, giving them the possibility to construct individually tailored activities, and they felt motivated to carry out their rehabilitation tasks on their own initiative. Feeling free to plan one's own activities and move about as desired is generally associated with a greater feeling of autonomy. This is confirmed in another qualitative substudy in the Teledialog project that focused on patients' use of a digital pedometer [35]. We have not identified other studies that support this finding. Future research needs to explore this perspective in depth over time.

The TTP increased the degree of relatedness between patients, health care professionals, and patients' partners. Patients found that the health care professionals became coaches for them in integrating their rehabilitation activities into their everyday life. Similar results have been found in a Danish study on patients with chronic pulmonary diseases, where a community of practice developed between patients and health care professionals who also took on a role as coaches for the benefit of the patients in their rehabilitation [36].

Some patients felt that they missed being part of a community of practice with other cardiac patients. Indeed, some patients might prefer to participate in rehabilitation activities at a health care center, where they can have more face-to-face contact and can feel part of a vibrant community of fellow patients and staff. One study, for example, found that patients can be reluctant to use new telehealth services if they place a high value on the existing services [32].

A few patients stated that they did not use the ActiveHeart Web forum for communication between each other. They considered this kind of open forum communication to be too private for their taste. One possible factor in explaining patients' hesitancy to use the Web forum might be the age of the enrolled patients. Older persons tend to be less accustomed to using social media to communicate with each other. The Web forum did not seem to too important to our patients. We discovered that patients make choices about what kind of digital media they want to use. In our study, the patients made use of the other digital solutions. Hence, they could refuse to use the Web forum but still use other digital media or solutions. Users apparently make informed choices. The findings that patients did not feel part of a community and avoided using the Web forum indicate that a telerehabilitation approach might not be suitable to all cardiac patients.

When designing telerehabilitation programs and technologies, it is important to stimulate interactive relationships among the involved parties. Recognizing someone's efforts also acknowledges the relationship and its importance, thus fulfilling the need for relatedness [19]. In the TTP, both patients and partners had access to the same technologies (data in the e-rehabilitation plan and the ActiveHeart.dk Web portal), and these technologies that might have helped them to be better able to communicate and share more concerns, anxieties, and questions.

In the SDT, the competency factor denotes the need to feel competent when taking on a task or new behavior [19]. The patients in our study stated that by participating in the TTP, they were able to integrate rehabilitation activities into their everyday and working lives. Being able to integrate rehabilitation into everyday life also makes it easier to master the "work" of being a patient. When rehabilitation becomes an extra task that needs to be fitted into an already busy schedule, the motivation to keep it up may diminish. We know that working patients find it difficult to participate in rehabilitation because of scheduling, and sheer distance to a health center may be another factor affecting the level of adherence [3].

Results from a survey in The Teledialog project comparing the telerehabilitation and control group showed that both groups were equally motivated for lifestyle changes and self-care and that they experienced a similar level of quality of life [37].

The goal of any telerehabilitation program is to allow patients the temporal and logistical freedom to organize their own rehabilitation process, thus giving them a competency-based motivation. They can focus on their rehabilitation process after other tasks have been completed, and they do not need to consider the time and expense of traveling to a rehabilitation center. This flexibility may also decrease the patients' feeling of a disparity between being a patient and reintegrating into daily life. If coping with lifestyle changes and disease management are to be successful, the rehabilitation activities need to become part of people's everyday life; hence, telerehabilitation can facilitate this need for enhanced integration.

Our findings indicated that integrating a partner into a telerehabilitation process gave additional partner involvement and allowed partners to become more engaged than they could be in conventional rehabilitation. The additional involvement factor was attributed to the functionalities of the technologies used. Jansson et al [29] concluded that the involvement of patients, partners, and stakeholders led to person-centered care that included maintaining social relations and being able to maintain other activities, including regular work. Person-centered care or treatment trends are certainly going to become a more common health care strategy in the future [38].

Both patients and partners expressed a sense of security related to taking part in the telerehabilitation program. At the same time, patients' partners acknowledged the challenges of navigating between being involved, being an equal partner, or not pushing the patient too hard. The enhanced sense of security felt by both patients and their partners was also a positive finding of our study, in view of the fact that overprotection is a common problem in couples dealing with cardiac disease [14]. Overprotective or controlling behavior by the cohabiting partner often diminishes the couple's ability to cope communally with the challenges of rehabilitation [14]. An overprotective or controlling partner may reflect their anxiety about the patient falling ill again or worsening of the disease. The ability to release some of the understandable anxiety was related to a partner's serious disease and having one's questions answered, and concerns addressed by health care staff make it possible for partners to engage more actively in the relationship and to assist their partner in facing rehabilitation challenges. Hence, the couple relationship, instead of being a source of tension, may serve as a source of social support for both the patient and partner. Social support has been strongly associated with better outcomes. However, the degree of positive social support is dependent on the quality of the relationship [15]. We must also be aware that partners' tendency toward controlling and overprotective behavior can be reduced by recognizing their important role in rehabilitation and by involving them in the rehabilitation process from the earliest stages.

Limitations

Case studies that are properly carried out can provide grounds for generalization [39]. First, in order to ensure the validity of our case study, a triangulation of data collection techniques was used. Interviews with the patients and partners took place at the same time. It is certainly possible that the presence of patients together with their partners might have influenced the conversation and the degree of respondents' openness during the interviews. However, we chose to conduct the interviews in pairs in order to have openness in the interview situation.

Second, we used a software program to analyze data. This kind of analysis may decontextualize the data on the rehabilitation sequence, leading to loss of valuable insights into what is certainly a complex process.

We are aware that the results and discussion on telerehabilitation on the motivation of cardiac patients do not allow us to fully discuss this aspect in this qualitative study. However, we have explored this issue in a survey in the Teledialog study [37].

Conclusions

Cardiac patients and their partners found telerehabilitation technologies a helpful digital toolbox in the rehabilitation process. Telerehabilitation motivated the patients to integrate rehabilitation activities into their work schedule and everyday life and made them feel like unique individuals. Participating in a telerehabilitation program might not be a suitable strategy for all cardiac patients. Being a patient's partner in a telerehabilitation program was associated with a heightened sense of security about the process, and the partners had to navigate between their involvement in the process, being an equal partner, and not pushing the patient too hard. Future research needs to focus on longitudinal case studies of both patients' and their partners' motivation in order to determine the benefits of participation in a telerehabilitation programs.

Acknowledgments

We wish to thank the patients and clinical partners who participated in our project, as well as the partners in the project who have invested their time. The Teledialog research project was funded by Eir Research and Business Park, The European Regional Development Fund, UNIK Partnership, Danish Council of Research and Innovation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Teledialog video.

[PNG File, 1MB - jmir_v21i4e13281_app1.png]

Multimedia Appendix 2

Observation guide.

[DOCX File, 14KB - jmir_v21i4e13281_app2.docx]

Multimedia Appendix 3

Interview guide for patients and their partners.

[DOCX File, 28KB - jmir_v21i4e13281_app3.docx]

References

1. World Heath Organization. 2019. Cardiovascular disease URL: https://www.who.int/cardiovascular_diseases/en/ [accessed 2019-02-18] [WebCite Cache ID 76H29tsJL]
2. Blair J, Corrigall H, Angus NJ, Thompson DR, Leslie S. Home versus hospital-based cardiac rehabilitation: a systematic review. Rural Remote Health 2011;11(2):1532 [FREE Full text] [Medline: 21488706]
3. Würgler MW, Sonne LT, Kilsmark J, Voss H, Søgaard J. Danish heart patients' participation in and experience with rehabilitation. Scand J Public Health 2012 Mar;40(2):126-132. [doi: 10.1177/1403494811435487] [Medline: 22316573]
4. Dalal HM, Doherty P, Taylor RS. Cardiac rehabilitation. BMJ 2015 Sep 29;351:h5000 [FREE Full text] [Medline: 26419744]
5. Haesum L, Nielsen G, Hangaard SV, Hejlesen O, Dinesen B. Identification of heart patients' everyday need: Based on user-driven innovation. In: Global Telemedicine and eHealth Updates: Knowledge Resources. Luxembourg: International Society for Telemedicine & eHealth; 2012 Presented at: Med@Tel; April 18-20, 2012; Luxembourg p. 611-614.
6. Conraads VM, Deaton C, Piotrowicz E, Santaularia N, Tierney S, Piepoli MF, et al. Adherence of heart failure patients to exercise: barriers and possible solutions: a position statement of the Study Group on Exercise Training in Heart Failure of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2012 May;14(5):451-458 [FREE Full text] [doi: 10.1093/eurjhf/hfs048] [Medline: 22499542]
7. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]
8. Brennan D, Tindall L, Theodoros D, Brown J, Campbell M, Christiana D, et al. A blueprint for telerehabilitation guidelines. Int J Telerehabil 2010;2(2):31-34 [FREE Full text] [doi: 10.5195/ijt.2010.6063] [Medline: 25945175]
9. Munro J, Angus N, Leslie SJ. Patient focused Internet-based approaches to cardiovascular rehabilitation--a systematic review. J Telemed Telecare 2013 Sep;19(6):347-353. [doi: 10.1177/1357633X13501763] [Medline: 24163299]
10. Hailey D, Roine R, Ohinmaa A, Dennett L. Evidence of benefit from telerehabilitation in routine care: a systematic review. J Telemed Telecare 2011;17(6):281-287. [doi: 10.1258/jtt.2011.101208] [Medline: 21844172]
11. Piotrowicz E, Piotrowicz R. Cardiac telerehabilitation: current situation and future challenges. Eur J Prev Cardiol 2013 Jun;20(2 Suppl):12-16. [doi: 10.1177/2047487313487483c] [Medline: 23702985]
12. Chan C, Yamabayashi C, Syed N, Kirkham A, Camp PG. Exercise Telemonitoring and Telerehabilitation Compared with Traditional Cardiac and Pulmonary Rehabilitation: A Systematic Review and Meta-Analysis. Physiother Can 2016;68(3):242-251. [doi: 10.3138/ptc.2015-33] [Medline: 27909373]
13. Cristo D, Nascimento N, Dias A, Sachetti A. Telerehabilitation for cardiac patients: systematic review. Int J Cardiovasc Sci 2018;31:443-450. [doi: 10.5935/2359-4802.20180032]
14. Dalteg T, Benzein E, Fridlund B, Malm D. Cardiac disease and its consequences on the partner relationship: a systematic review. Eur J Cardiovasc Nurs 2011 Sep;10(3):140-149. [doi: 10.1016/j.ejcnurse.2011.01.006] [Medline: 21345737]
15. Nissen NK, Jónsdóttir M, Spindler H, Zwisler AO. Resistance to change: Role of relationship and communal coping for coronary heart disease patients and their partners in making lifestyle changes. Scand J Public Health 2018 Aug;46(6):659-666. [doi: 10.1177/1403494818756562] [Medline: 29400128]
16. Hippel VE. Democratizing Innovation. Cambridge, MA: The MIT Press; 2019.
17. Wenger E, White N, Smith JD. Digital Habitats: Stewarding Technology For Communities. Portland, Oregon: CPSQUARE; 2009.
18. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
19. Ryan R, Heather P, Deci E, Geoffrey W. The European Health Psychologist. 2008. Facilitating health behavior change and its maintenance: Interventions based on Self-Determination Theory URL: http://selfdeterminationtheory.org/SDT/documents/2008_RyanPatrickDeciWilliams_EHP.pdf [accessed 2019-01-01] [WebCite Cache ID 756o88yAV]
20. Yin R. Case Study Research: Design and Methods (Applied Social Research Methods). London: Sage Publications; 2014.
21. Sundhedsstyrelsen. 2004. Vejledning om hjerterehabilitering på sygehuse URL: https://www.sst.dk/~/media/094F8791F82B4CBC9166314D2F97AAA2.ashx [accessed 2019-01-01] [WebCite Cache ID 756oHzyFd]
22. Sundhedsstyrelsen. 2015. National klinisk retningslinje for hjerterehabilitering URL: https://www.sst.dk/da/udgivelser/2015/~/media/401919781C684EE9AAE544EB5E76847B.ashx [accessed 2019-01-01] [WebCite Cache ID 756odgrz7]
23. Piepoli MF, Corrà UGO, Benzer W, Bjarnason-Wehrens B, Dendale P, Gaita D, Cardiac Rehabilitation Section of the European Association of Cardiovascular PreventionRehabilitation. Secondary prevention through cardiac rehabilitation: from knowledge to implementation. A position paper from the Cardiac Rehabilitation Section of the European Association of Cardiovascular Prevention and Rehabilitation. Eur J Cardiovasc Prev Rehabil 2010 Feb;17(1):1-17. [doi: 10.1097/HJR.0b013e3283313592] [Medline: 19952757]

24. Uchino B, Bowen K, Carlisle MCM, Birmingham W. Psychological pathways linking social support to health outcomes: a visit with the "ghosts" of research past, present, and future. Soc Sci Med 2012 Apr;74(7):949-957 [FREE Full text] [doi: 10.1016/j.socscimed.2011.11.023] [Medline: 22326104]
25. Kristiansen S. Deltagende Observation: Introduktion Til En Samfundsvidenskabelig Metode. Copenhagen: H Reitzels; 2004.
26. Kvale S, Brinkmann S. Interviews: Learning The Craft Of Qualitative Research Interviewing. Thousand Oaks, CA: Sage Publications; 2008.
27. Timmins F. A review of the information needs of patients with acute coronary syndromes. Nurs Crit Care 2005;10(4):174-183. [Medline: 15997971]
28. Dinesen B, Spindler H. The Use of Telerehabilitation Technologies for Cardiac Patients to Improve Rehabilitation Activities and Unify Organizations: Qualitative Study. JMIR Rehabil Assist Technol 2018 Nov 19;5(2):e10758 [FREE Full text] [doi: 10.2196/10758] [Medline: 30455168]
29. Jansson I, Fors A, Ekman I, Ulin K. Documentation of person-centred health plans for patients with acute coronary syndrome. Eur J Cardiovasc Nurs 2018 Dec;17(2):114-122. [doi: 10.1177/1474515117718415] [Medline: 28657364]
30. Brouwers RWM, Kraal JJ, Traa SCJ, Spee RF, Oostveen LMLC, Kemps HMC. Effects of cardiac telerehabilitation in patients with coronary artery disease using a personalised patient-centred web application: protocol for the SmartCare-CAD randomised controlled trial. BMC Cardiovasc Disord 2017 Dec 31;17(1):46 [FREE Full text] [doi: 10.1186/s12872-017-0477-6] [Medline: 28143388]
31. 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 Dec;101(5):854-861 [FREE Full text] [doi: 10.1016/j.pec.2017.12.017] [Medline: 29305064]
32. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]
33. Anttila M, Kivistö H, Piirainen A, Kokko K, Malinen A, Pekkonen M, et al. Cardiac Rehabilitees' Technology Experiences Before Remote Rehabilitation: Qualitative Study Using a Grounded Theory Approach. J Med Internet Res 2019 Feb 07;21(2):e10985 [FREE Full text] [doi: 10.2196/10985] [Medline: 30730298]
34. Albert NM, Dinesen B, Spindler H, Southard J, Bena JF, Catz S, et al. Factors associated with telemonitoring use among patients with chronic heart failure. J Telemed Telecare 2017 Feb;23(2):283-291. [doi: 10.1177/1357633X16630444] [Medline: 26869144]
35. Thorup CB, Grønkjær M, Spindler H, Andreasen JJ, Hansen J, Dinesen BI, et al. Pedometer use and self-determined motivation for walking in a cardiac telerehabilitation program: a qualitative study. BMC Sports Sci Med Rehabil 2016;8:24 [FREE Full text] [doi: 10.1186/s13102-016-0048-7] [Medline: 27547404]
36. Dinesen B, Huniche L, Toft E. Attitudes of COPD patients towards tele-rehabilitation: a cross-sector case study. Int J Environ Res Public Health 2013 Nov 15;10(11):6184-6198 [FREE Full text] [doi: 10.3390/ijerph10116184] [Medline: 24247995]
37. Spindler H, Leerskov K, Joensson K, Nielsen G, Andreasen JJ, Dinesen B. Conventional Rehabilitation Therapy Versus Telerehabilitation in Cardiac Patients: A Comparison of Motivation, Psychological Distress, and Quality of Life. Int J Environ Res Public Health 2019 Feb 12;16(3):1-15 [FREE Full text] [doi: 10.3390/ijerph16030512] [Medline: 30759761]
38. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized Telehealth in the Future: A Global Research Agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
39. Seale C, Gobo G, Silverman D. Five misunderstandings about case-study research. In: Qualitative Research Practice: Concise Paperback Edition. London: Sage Publications; 2019.

Abbreviations

N/A: not applicable
SDT: Self-determination Theory
TTP: Teledialog Telerehabilitation Program

Edited by G Eysenbach; submitted 02.01.19; peer-reviewed by C Karlsen, R Brouwers, EK Ghandour, A Lee; comments to author 24.01.19; revised version received 21.02.19; accepted 24.03.19; published 15.04.19.

Please cite as:
Dinesen B, Nielsen G, Andreasen JJ, Spindler H
Integration of Rehabilitation Activities Into Everyday Life Through Telerehabilitation: Qualitative Study of Cardiac Patients and Their Partners
J Med Internet Res 2019;21(4):e13281
URL: http://www.jmir.org/2019/4/e13281/
doi:10.2196/13281
PMID:30985284

Original Paper

Efficacy of a Web-Based Self-Management Enhancing Program for Patients with Rheumatoid Arthritis: Explorative Randomized Controlled Trial

Rixt Zuidema[1], MSc; Sandra van Dulmen[2,3,4], PhD; Maria Nijhuis-van der Sanden[1], PhD; Inger Meek[5], PhD; Cornelia van den Ende[6], PhD; Jaap Fransen[5], PhD; Betsie van Gaal[1,7], RN, PhD

[1]Scientific Institute for Quality of Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[2]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[3]Netherlands Institute for Health Services Research, Utrecht, Netherlands
[4]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway
[5]Department of Rheumatology, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[6]Department of Rheumatology and Pharmacy, Sint Maartenskliniek, Nijmegen, Netherlands
[7]Institute of Nursing, Hogeschool van Arnhem en Nijmegen University of Applied Sciences, Nijmegen, Netherlands

Corresponding Author:
Rixt Zuidema, MSc
Scientific Institute for Quality of Healthcare
Radboud Institute for Health Sciences
Radboud University Medical Center
Postbus 9101, Huispost 114
Nijmegen, PO Box 9101, 6500
Netherlands
Phone: 31 2436 5567
Email: rixt.zuidema@radboudumc.nl

Abstract

Background: Web-based self-management enhancing programs have the potential to support patients with rheumatoid arthritis (RA) in their self-management; for example, improve their health status by increasing their self-efficacy or taking their prescribed medication. We developed a Web-based self-management enhancing program in collaboration with RA patients and professionals as co-designers on the basis of the intervention mapping framework. Although self-management programs are complex interventions, it is informative to perform an explorative randomized controlled trial (RCT) before embarking on a larger trial.

Objective: This study aimed to evaluate the efficacy of a Web-based self-management enhancing program for patients with RA and identify outcome measures most likely to capture potential benefits.

Methods: A multicenter exploratory RCT was performed with an intervention group and a control group. Both groups received care as usual. In addition, the intervention group received 12 months of access to a Web-based self-management program. Assessment occurred at baseline, 6 months, and 12 months. Outcome measures included self-management behavior (Patient Activation Measurement, Self-Management Ability Scale), self-efficacy (Rheumatoid Arthritis task-specific Self-Efficacy, Perceived Efficacy in Patient-Physician Interaction), general health status (RAND-36), focus on fatigue (Modified Pain Coping Inventory for Fatigue), and perceived pain and fatigue (Numeric Rating Scales). A linear mixed model for repeated measures, using the intention-to-treat principle, was applied to study differences between the patients in the intervention (n=78) and control (n=79) groups. A sensitivity analysis was performed in the intervention group to study the influence of patients with high (N=30) and low (N=40) use of the intervention.

Results: No positive effects were found regarding the outcome measurements. Effect sizes were low.

Conclusions: Based on these results, it is not possible to conclude on the positive effects of the intervention or to select outcome measures to be regarded as the primary/main or secondary outcomes for a future trial. A process evaluation should be performed to provide more insight into the low compliance with and effectiveness of the intervention. This can determine for whom this sort of program will work and help to fine-tune the inclusion criteria.

Trial Registration: Netherlands Trial Register NTR4871; https://www.trialregister.nl/trial/4726

(J Med Internet Res 2019;21(4):e12463) doi:10.2196/12463

KEYWORDS

self-management; internet; arthritis, rheumatoid

Introduction

Background

Rheumatoid arthritis (RA) is one of the most prevalent chronic conditions, with a pervasive impact on daily life [1]. Despite the introduction of biological therapies and conventional disease-modifying antirheumatic drugs, RA patients experience a high level of pain [2] and fatigue [3,4], which leads to disabilities like restrictions in work participation [5,6] and leisure activities [7-9]. Moreover, many RA patients experience disease-related psychological problems, like depressive mood and helplessness [10,11].

To optimally manage the consequences of RA and reduce the impact of the disease on patients in daily life, effective self-management programs are needed. Web-based self-management programs can easily reach a large group of RA patients in their own place and time and provide more anonymity than face-to-face programs. Studies have shown that patients feel more comfortable sharing sensitive information like reports on daily activity or feelings online [12]. Other advantages are the possibility of tailoring information, avoiding waiting lists, and 24-hour availability [13].

Studies about Web-based self-management programs have shown to be effective in RA patients on several health outcomes, including increased self-efficacy, knowledge and physical activity [14], less pain, disability and depression, and reduction in the overuse of medication and the number of visits to physicians [15-17]. However, many of the programs are developed without end-user involvement. Consequently, these programs may not suit patient support needs for self-management as patient preferences for program use are not well known [18,19].

To guarantee optimal patient involvement, we developed a Web-based self-management program on the basis of intervention mapping (IM), called Reuma zelf te lijf (Coping with RA) [20-22]. According to the Medical Research Council (MRC), complex interventions such as this program can be evaluated in a randomized controlled trial (RCT); however, it is advised to first perform an explorative study investigating potential outcome measurements to be used in a larger trial [23].

Objectives

Therefore, the present explorative RCT study in patients with RA was aimed as follows: (1) to explore the potential efficacy of a Web-based self-management enhancing program versus “usual care” on self-management behavior, self-efficacy, general health status, coping with fatigue and the level of pain and fatigue and to determine the effect sizes at 6 and 12 months after baseline, and (2) to identify outcome measures most likely to capture the potential benefits covered by the performance objectives, by exploring their floor and ceiling effects at baseline.

Methods

Design

A multicenter exploratory RCT was conducted in 2 Dutch hospitals, The Radboudumc (a University hospital) and the Sint Maartenskliniek (a specialized hospital in rheumatology, rehabilitation, and orthopedic surgery), both located in Nijmegen, the Netherlands. An intervention and a control group were compared at 6 and 12 months after baseline on 6 outcome measurements to explore the efficacy of the Web-based program and to identify outcome measures [20]. The trial is registered at the Netherlands Trial Register (ID: NTR4871).

Ethical Approval

The medical ethics committee of Arnhem-Nijmegen approved this study (No. 2014-1208).

Participants

Between December 2014 and June 2015, patients with a diagnosis of RA aged 18 years or older were invited by a letter to participate in this study, in collaboration with rheumatologists, until the required number of 190 patients was reached. Patients received the following: (1) information about the study, (2) a questionnaire for screening eligibility, and (3) an informed consent form. Eligibility criteria were the ability to speak and read Dutch and having access to a computer with an internet connection. Patients receiving psychiatric or psychological treatment were excluded. Patients willing to participate were asked to return the informed consent with the completed questionnaire. When patients agreed to participate and were eligible, the researcher sent the patient an email with the baseline questionnaire.

Randomization

Eligible patients were stratified by the hospital and randomly assigned to the intervention or control group by an independent statistician using an automated randomization program. The researcher informed the patients by post if they were allocated to the control or intervention group. Patients in the control group continued with their care as usual, which comprised medical treatment at the outpatient clinic. The patients in the intervention group received, in additional to their care as usual, 12 months of access to the intervention directly after randomization.

Table 1. Overview of the 9 modules and their performance objectives.

Module name	Performance objective: Patients need to...
Balancing activity and rest	...find balance between rest and activity; make choices when participating in daily life activities to keep balance
Setting boundaries	...set boundaries for their partner, relatives, colleagues and social environment
Asking for help and social support	...ask for social support or practical help from their partner, relatives, colleagues and social environment in daily life; ask for social support and practical help from colleagues; accept receiving social support or practical help from their partner, relatives, colleagues and social environment in daily life
Use of medicines	...take prescribed medication
Communication with health professionals	...prepare for a visit to a health professional; ask questions and/or express concerns during an appointment with a health professional
Use of assistive devices	...use, if necessary, assistive devices
Performing physical exercises	...perform daily physical exercises
Coping with worries	...cope with worries about RA[a]
Coping with RA	...cope with RA

[a] RA: rheumatoid arthritis.

Intervention

Web-Based Self-Management Enhancing Program

The intervention was developed between January 2013 and July 2014 in collaboration with RA patients and professionals as co-designers [20,24]. The theory of planned behavior was used as the underlying theory and essential behavioral change techniques were applied to induce behavioral change formulated as performance objectives, selected according to the IM steps [21,25,26]. The Web-based self-management enhancing program comprises 9 modules with 13 performance objectives (Table 1) and a diary to track patients' fatigue and pain over time [20]. Each module comprises 2-5 sessions, with informational and persuasive texts, videos with instructions and role models, exercises, and assignments. The program is unguided, and patients need to choose a module by their own and can work through it at their own pace whenever they want.

Implementation of the Web-Based Self-Management Enhancing Program

To implement the Web-based program and to increase use of the program by patients, 3 implementation strategies were deployed during the study: (1) patients received a written instruction manual for the program, (2) reminders to (re)visit the program were sent twice weekly via email, and (3) nurses brought the program to the attention of the intervention group participants during their consultation.

Measurements and Outcomes

All included patients who filled in the baseline questionnaire between January 2015 and June 2015 received a questionnaire after 6 months (T1) and 12 months (T2). At baseline, demographic and disease characteristics were assessed. Patient-reported outcome measurements were assessed at baseline and during follow-ups (T1 and T2). When patients preferred a paper questionnaire, a version was sent by post. When patients did not return the questionnaire at T1 but filled in the questionnaire at T2, this was indicated as a missing value at T1. Patients who did not return the T2 questionnaire are indicated as dropouts.

Baseline Characteristics

The following demographic and disease characteristics were assessed: age, gender, education level, employment status, disease duration, Numeric Rating Scales (NRS) pain/fatigue, Modified Health Assessment Questionnaire (M-HAQ) physical disability, and satisfaction with health status. The M-HAQ comprises 8 questions on difficulties in daily activities in the following domains: dressing, rising, eating, walking, hygiene, reaching, gripping, and usual activities. Patients responded on a 4-point scale, with a higher score indicating more difficulty in performing daily activities. Health satisfaction was assessed using 1 question about patients' (dis)satisfaction about the course of their disease last week, with 4 response options, with a higher score meaning less satisfied than before and an *"I don't know"* option [27].

Outcome Measurements

Based on the theory of planned behavior, 6 outcome measures were relevant: self-management behavior, self-efficacy, general health status, coping with fatigue, and the level of pain and fatigue.

Self-Management Behavior

The Patient Activation Measurement (PAM-13) includes statements about individuals' knowledge, confidence, and skills for self-management of their chronic illness behavior and the level of activation. It includes 13 items on a 5-point scale with a higher score indicating a higher level of patient activation. The scores of the 13 items are summarized as a total score. Total PAM scores were computed if at least 10 items were completed [28].

The short Self-Management Ability Scale (SMAS-S) comprises 18 items scored on a 6-point scale with a higher score indicating better self-management behavior [29].

Self-Efficacy

The Rheumatoid Arthritis task-specific Self-Efficacy (RASE) questionnaire comprises 28 items scored on a 5-point Likert scale. Higher scores reflect higher self-efficacy [30]. This questionnaire was translated into Dutch via forward-backward translation and decisions were based on consensus with a group of 5 researchers, 4 RA patients, and 1 RA patient who was a native English speaker. The Perceived Efficacy in Patient-Physician Interaction (PEPPI-5) comprises a 5-point Likert scale. A higher score reflects more confidence in patient interactions with their physician [31].

General Health Status

The RAND-36 comprises 36 questions measuring 8 dimensions: physical functioning, social functioning, physical role limitations, emotional role limitations, mental health, vitality, pain, perceived health-related quality of life and behavioral change, with various response options based on 3- to 6-point Likert scales, with a higher score indicating better perceived health-related quality of life. Scores were transformed to a 0-100-point scale for each subscale [32].

Level of Pain and Fatigue

Pain and fatigue were measured with NRS, ranging from 0 to 10 with 0 indicating no pain/fatigue and 10 indicating severe pain/very tired. For both outcomes, 2 questions were asked: the level of pain/fatigue today and the mean level of pain/fatigue during the last 2 weeks.

Coping with Fatigue

The Modified Pain Coping Inventory for Fatigue (MPCI-F) was used. This questionnaire is based on a subscale of the Pain Coping Inventory questionnaire and modified to assess coping with fatigue instead of coping with pain [33]. The questionnaire comprises 8 items to assess the focus on fatigue. A higher score reflects more focus on fatigue.

Statistical Analysis

Descriptive statistics were used to describe the control and intervention groups at baseline. *t* tests and Chi-square tests were used to analyze baseline differences. It was analyzed whether the patients who dropped out differed from the group that returned the questionnaire at T2 [34]. Between-group differences in outcomes were analyzed using a linear mixed model to account for repeated measurements and to handle missing data under the missing-at-random assumption. Differences between the intervention and control groups were analyzed at baseline, after 6 months (T1), and 12 months (T2). The fixed variables in the model were as follows: group (intervention/control), hospital (hospital 1 or hospital 2), age, gender, disease duration, education level, employment status, physical functioning (M-HAQ), and the interaction terms between measurement time points and groups. The first analysis was done using the intention-to-treat principle. Subsequently, a sensitivity analysis was performed to explore the influence of program use within the intervention group. The intervention group was divided into 3 groups: (1) a group with low usage (0-1 visits), (2) a group with moderate usage (2-5 visits), and (3) a group with high usage (6 or more visits). In the analysis, the group with moderate usage was left out to increase the contrast between the groups with low and high usage. *t* tests and Chi-square tests were performed to analyze between-group differences in demographics, disease-related characteristics, and outcomes at baseline, T1, and T2. Statistical significance was defined as $P<.05$.

For all outcome measurements, Cohen *d* was used to quantify effect sizes by calculating the difference in means, divided by the pooled within-group standard deviation [35]. Following Cohen definition of effect sizes, less than 0.4 was defined as a small effect, between 0.5 and 0.7 as moderate, and ≥0.8 was considered as a large effect [36]. Floor and ceiling effects were explored for all outcome measures by examining the percentage of minimum and maximum scores, which reflects the extent that patients scored the lowest or the highest score. For a 3- or 5-point Likert scale, floor and ceiling effects were defined as more than 80% of the patients scoring the lowest/highest. Statistical analyses were performed using SPSS version 22 (SPSS Inc) for Windows. For exploratory RCT such as these, sample sizes are not calculated based on formal power analyses. For this trial, a sample size of 200 patients was chosen, which was considered a sufficient size for a representation of the relevant variation in the target group.

Results

Overview

In total, 669 patients were eligible and invited. Of these, 191 patients expressed interest and 189 met the inclusion criteria (see Figure 1). In total, 157 patients completed the baseline questionnaire between January 2015 and June 2015. These patients were randomly assigned to the intervention group (n=78) and the control group (n=79), stratified by hospital. At T1, 59 in the intervention group and 65 in the control group filled in the questionnaire. At T2, 54 patients in the intervention group and 74 patients in the control group completed the questionnaire. Overall, in the intervention group, few patients (69%, 54/78) participated at T2 than in the control group (94%, 74/79). Most of these patients gave the burden of their illness as the reason for dropout. Some patients refused to fill in the questionnaire at T1 but completed the questionnaire at T2, which explains the higher number of patients who filled in the questionnaire at T2 compared with T1. Differences in demographics and disease-related characteristics between the group of patients who refused to fill in the questionnaire at T2 and the group who returned the questionnaire at T2 were small (<10%), which indicated that dropout did not influence the outcomes.

Figure 1. CONSORT flow diagram.

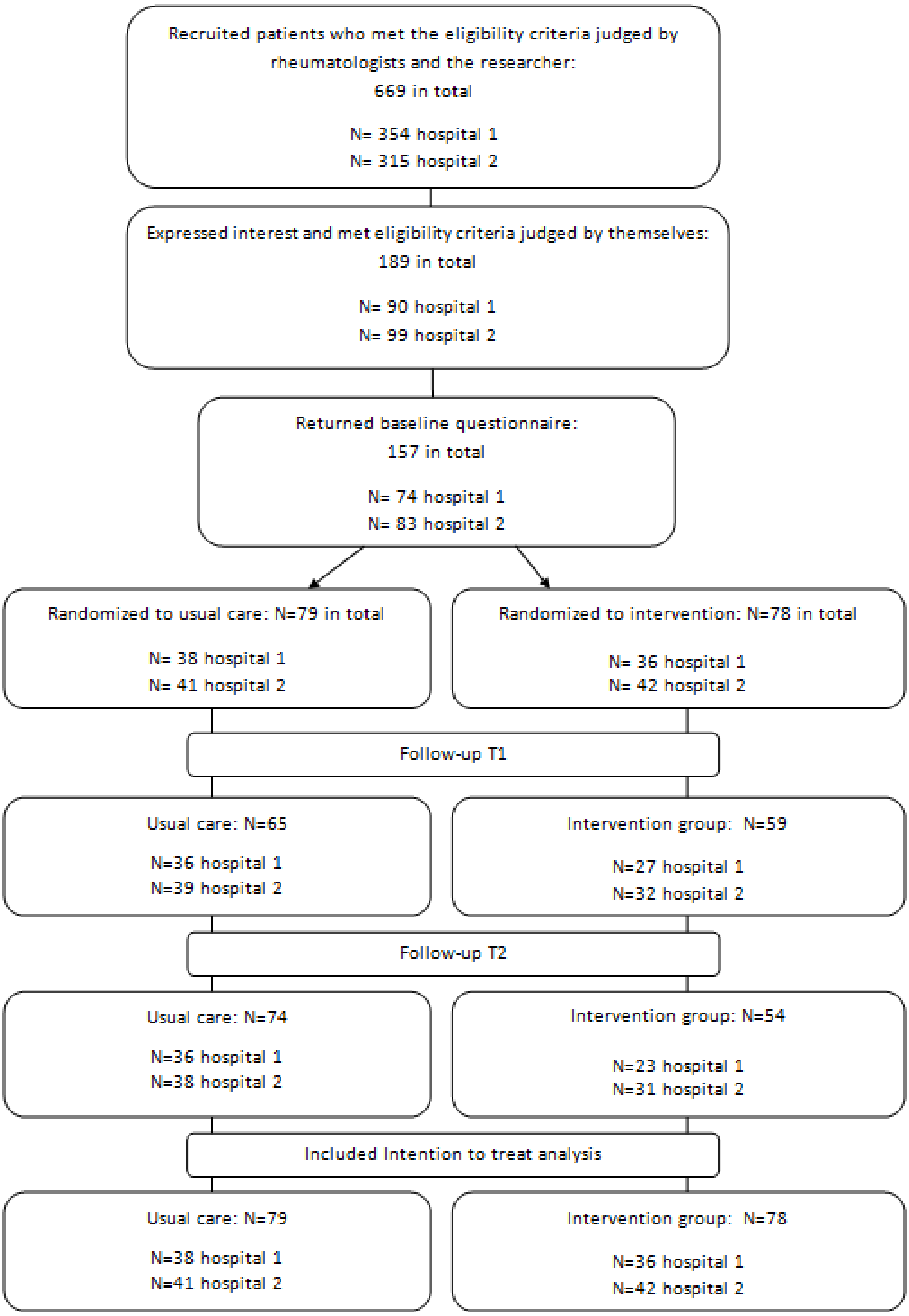


Table 2. Demographics and disease-related characteristics at baseline.

Characteristics	Control group		Intervention group	
	N	Statistics	N	Statistics
Age (years), mean (SD)		62.9 (10.2)		61.0 (11.3)
Gender, %				
Men	27	34	27	35
Women	52	66	51	65
Disease duration, median (25th, 75th percentiles)	79	17 (6.0, 26)	77	9 (5.0, 19.5)
Education level, %				
Low	28	35	10	13
Medium	28	35	43	55
High	23	29	25	32
Employment status, %				
Not working	50	63	41	53
Part-time working	7	9	7	9
Working	22	28	30	39
Physical disability (M-HAQ[a]), median (25th, 75th percentiles)	79	0.5 (0.1, 1.4)	78	0.6 (0.1, 1.1)
NRS pain today, mean (SD)	79	3.3 (2.3)	77	3.2 (2.2)
NRS mean pain last 2 weeks, mean (SD)	79	3.9 (2.3)	78	3.6 (2.3)
NRS fatigue today (NRS), mean (SD)	79	4.1 (2.5)	78	3.8 (2.4)
NRS mean fatigue last 2 weeks, mean (SD)	79	4.3 (2.4)	78	4.3 (2.3)

[a] M-HAQ: Modified Health Assessment Questionnaire.

Baseline Characteristics of Patients

Demographics and disease-related characteristics at baseline were compared for the control group and the intervention group, as shown in Table 2. The only significant between-group difference in the patient characteristics was education level (*P*=.003). Fewer patients in the intervention group had a lower education level (12.8% vs 35.4%) and more patients had a moderate (55.1% vs 35.4%) or higher education level (32.1% vs 29.1%). Some patients who filled in a paper questionnaire did not complete all items, which explains the missing data in Tables 3 and 4.

The Outcome Measurements at Baseline and Follow-up

Table 3 gives an overview of the mean scores of outcome measurements of the patients in the intervention and control groups at baseline and after 6 and 12 months. The baseline scores of the 2 groups did not differ significantly.

Table 3. Mean scores of outcome measurements on baseline, T1 and T2 of control and intervention groups.

Scales, group	Baseline (T0)		T1		T2	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
PAM[a] (10-65)						
Control	57	46.9 (4.9)	49	47.7 (4.8)	45	47.8 (3.8)
Intervention	47	47.2 (3.7)	35	46.7 (6.9)	31	47.8 (2.9)
SMAS-S[b] (0-60)						
Control	79	36.0 (6.3)	75	37.9 (6.8)	74	37.6 (6.8)
Intervention	78	36.7 (7.1)	57	39.4 (6.4)	54	38.8 (7.0)
RASE[c] (28-140)						
Control	79	99.4 (12.7)	75	101.5 (10.6)	74	99.9 (11.6)
Intervention	78	102.9 (10.2)	57	101.9 (10.3)	54	102.0 (7.4)
PEPPI-5[d] (5-25)						
Control	79	21.6 (3.0)	75	21.0 (3.2)	73	20.6 (3.4)
Intervention	78	21.2 (3.3)	57	21.3 (3.1)	54	20.8 (3.1)
RAND physical functioning (0-100)						
Control	78	58.1 (27.0)	75	59.4 (26.5)	74	61.8 (25.9)
Intervention	77	61.7 (26.1)	57	65.9 (27.3)	54	65.9 (26.7)
RAND social functioning (0-100)						
Control	79	73.3 (2476)	75	72.7 (22.3)	74	73.1 (22.4)
Intervention	78	71.3 (20.8)	57	77.0 (19.6)	54	70.8 (24.3)
RAND physical role limitations (0-100)						
Control	79	49.1 (43.6)	75	51.11 (45.3)	74	49.0 (43.1)
Intervention	78	49.0 (43.3)	56	57.9 (42.0)	54	49.1 (44.2)
RAND emotional role limitations (0-100)						
Control	79	75.1 (40.5)	73	84.9 (35.2)	74	78.8 (39.2)
Intervention	77	80.1 (36.4)	54	85.2 (31.5)	54	78.4 (37.3)
RAND mental health (0-100)						
Control	78	54.7 (14.3)	75	72.6 (16.7)	74	76.1 (14.6)
Intervention	78	52.6 (13.4)	56	76.5 (12.0)	54	75.9 (13.8)
RAND vitality (0-100)						
Control	78	51.2 (22.7)	75	53.9 (21.6)	74	56.3 (21.2)
Intervention	78	53.1 (19.4)	56	61.2 (15.1)	54	62.5 (14.5)
RAND pain (0-100)						
Control	79	59.9 (21.3)	75	60.8 (22.2)	74	66.1 (21.8)
Intervention	78	64.3 (22.3)	57	67.1 (21.0)	54	63.9 (22.1)
RAND general health perception (0-100)						
Control	79	52.5 (18.7)	75	47.8 (18.3)	72	48.1 (17.5)
Intervention	77	52.7 (20.8)	57	52.3 (19.5)	48	50.4 (19.1)
RAND health change (0-100)						
Control	79	44.9 (21.2)	75	50.3 (19.9)	74	44.3 (19.7)
Intervention	78	47.8 (23.2)	57	51.3 (20.8)	54	43.5 (23.9)
NRS[e] pain today (0-10)						

Scales, group	Baseline (T0)		T1		T2	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Control	79	3.3 (2.3)	75	3.2 (2.2)	72	3.0 (2.2)
Intervention	77	3.2 (2.2)	57	3.0 (2.3)	48	3.3 (2.3)
NRS mean pain last 2 weeks (0-10)						
Control	79	3.9 (2.3)	75	3.8 (2.1)	72	3.6 (2.2)
Intervention	78	3.6 (2.3)	57	3.4 (2.3)	48	3.9 (2.4)
NRS fatigue today (0-10)						
Control	79	4.1 (2.5)	75	3.8 (2.6)	72	3.7 (2.3)
Intervention	78	3.8 (2.4)	57	3.4 (2.4)	48	3.6 (2.4)
NRS mean fatigue last 2 weeks (0-10)						
Control	79	4.3 (2.4)	75	4.2 (2.6)	72	4.2 (2.4)
Intervention	78	4.3 (2.3)	57	3.7 (2.1)	48	4.0 (2.2)
MPCI-F[f] (4-32)						
Control	79	14.1 (4.8)	75	14.1 (4.7)	74	13.6 (4.3)
Intervention	78	14.1 (3.9)	57	13.3 (3.3)	54	13.6 (3.2)

[a] PAM: Patient Activation Measurement.

[b] SMAS-S: short Self-Management Ability Scale.

[c] RASE: Rheumatoid Arthritis Self-Efficacy.

[d] PEPPI-5: Perceived Efficacy in Patient-Physician Interaction.

[e] NRS: numerical rating scale.

[f] MPCI-F: Modified Pain Coping Inventory for Fatigue.

In Table 4, the estimated differences between the intervention and control groups of the intention-to-treat analysis at 6 and 12 months are presented. Overall, the scores show no significant differences and small effect sizes. Only the outcome measurement of the subscale RAND-36 vitality at T2 (5.41 95% CI 0.16-10.65, *P*=.04) showed a significant difference with respect to T0, with a small effect size (Cohen *d*) of 0.01 in favor of the intervention group. Floor and ceiling effects were explored for all specified outcomes at baseline but were not found.

Sensitivity Analysis

Baseline Characteristics of Patients

High users of the intervention scored statistically significantly better than low users of the intervention on the following baseline characteristics: physical disability (M-HAQ; *P*=.03), RAND-36 subscale social functioning (*P*=.02), RAND-36 subscale physical role limitations (*P*=.03), RAND-36 pain (*P*=.03), and all the NRS scales, that is, pain today (*P*=.002), mean pain last 2 weeks (*P*=.02), fatigue today (*P*<.001) and mean fatigue last 2 weeks (*P*<.001; see Table 5).

After performing the sensitivity analysis, a statistically significant effect was found for the group with high usage on the subscale RAND-36 general health perception after 12 months (9.65, 95% CI 0.83-18.48, *P*=.03), with a small effect size of 0.02 (Table 6). No floor and ceiling effects were found for any of the specified outcomes at baseline in the groups with low or high usage.

Table 4. The estimated group differences between intervention and control groups after intention-to-treat analysis at 6 months and 12 months after baseline.

Scales	6 months after baseline				12 months after baseline			
	T0-T1 change	95% CI	*P* value[a]	Cohen *d*	T0-T2 change	95% CI	*P* value[a]	Cohen *d*
PAM[b] (10-65)	–0.7	–3.4 to 1.5	.44	0.0	–0.1	–1.6 to 1.5	.93	0.00
SMAS-S[c] (0-60)	0.3	–1.4 to 2.0	.72	0.0	0.7	–1.1 to 2.5	.43	0.03
RASE[d] (28-140)	–2.1	–4.9 to 0.8	.16	0.0	0.3	–2.2 to 2.9	.81	0.00
PEPPI-5[e] (5-25)	0.4	–0.5 to 1.2	.40	0.0	0.3	–0.7 to 1.3	.51	0.03
RAND[f] physical functioning (0-100)	2.5	–3.3 to 8.1	.40	0.00	–0.2	–5.4 to 5.1	.96	0.00
RAND social functioning (0-100)	4.1	–1.5 to 9.6	.15	0.0	–2.7	–9.2 to 3.8	.42	–0.01
RAND physical role limitations (0-100)	5.6	–7.0 to 18.2	.38	0.00	–2.8	–14.9 to 9.3	.65	0.00
RAND emotional role limitations (0-100)	–3.2	–14.1 to 7.6	.56	0.00	–3.9	–16.0 to 8.3	.53	0.00
RAND mental health (0-100)	2.8	–1.1 to 6.8	.16	0.0	0.9	–3.0 to 4.7	.66	0.00
RAND vitality (0-100)	3.4	–1.5 to 8.3	.17	0.0	5.4	0.2 to 10.7	.04[g]	0.01
RAND pain (0-100)	2.6	–3.7 to 8.9	.42	0.0	–6.1	–12.5 to 0.4	.06	–0.01
RAND general health perception (0-100)	2.2	–2.2 to 6.7	.33	0.0	–0.1	–4.5 to 4.4	.98	0.00
RAND health change (0-100)	0.1	–6.8 to 7.1	.97	0.00	–1.4	–9.0 to 6.2	.72	0.00
NRS[h] pain today (0-10)	0.0	–0.6 to 0.7	.97	0.00	0.5	–0.1 to 1.2	.13	0.10
NRS mean pain last 2 weeks (0-10)	0.0	–0.7 to 0.6	.97	0.00	0.7	0.0 to 1.4	.60	1.13
NRS fatigue today (0-10)	0.2	–0.5 to 0.8	.66	0.0	0.3	–0.4 to 0.9	.46	0.01
NRS mean fatigue last 2 weeks (0-10)	–0.23	–0.9 to 0.4	.45	–0.1	0.1	–0.6 to 0.7	.81	0.00
MPCI-F[i] (4-32)	0.1	–0.8 to 0.9	.90	0.00	0.3	–0.7 to 1.2	.58	0.01

[a] Values represent outcomes of the ITT analysis without confounders. After adding confounders, no changes in values appear.

[b] PAM: Patient Activation Measurement.

[c] SMAS-S: short Self-Management Ability Scale.

[d] RASE: Rheumatoid Arthritis Self-Efficacy.

[e] PEPPI-5: Perceived Efficacy in Patient-Physician Interaction.

[f] RAND-36: General Health Status.

[g] Significant differences (*P*<.05) between control and intervention groups.

[h] NRS: numerical rating scale.

[i] MPCI-F: Modified Pain Coping Inventory for Fatigue.

Table 5. Scores at baseline for the groups with a low and high usage of the intervention: demographic characteristics, disease-related characteristics, and outcome measures.

Characteristics and outcome measures at baseline	Low usage		High usage		*P* value
	N	Statistics	N	Statistics	
Age (years), mean (SD)	29	63.8 (10.5)	40	58.9 (10.8)	.06
Gender, %					.55
Men	10	33	14	35	
Women	20	67	26	65	
Disease duration, (median (25th, 75th percentiles))	29	8,0 (4,5, 22,5)	40	8,5 (5,0, 18,7)	.80
Education level, %					.78
Low	3	10	5	13	
Middle	19	63	22	55	
High	8	28	13	33	
Employment status, %					.28
Not working	22	73	23	58	
Working	8	27	17	43	
Physical disability (Modified Health Assessment Questionnaire), (median [25th, 75th percentiles])	30	1.1 (0.2, 1.6)[a]	40	0.5 (0.1, 1,0)	.03[a]
PAM[b] (10-65), mean (SD)	20	48.0 (3.3)	20	46.2 (3.8)	.11
SMAS-S[c] (0-60), mean (SD)	30	36.5 (7.3)	40	37.7 (7.0)	.48
RASE[d] (28-140), mean (SD)	30	102.1 (10.9)	40	103.4 (9.1)	.58
PEPPI-5[e] (5-25), mean (SD)	30	21.5 (3.9)	40	21.2 (2.8)	.68
RAND[f] physical functioning (0-100), mean (SD)	29	54.3 (28.3)	40	66.3 (24.6)	.07
RAND social functioning (0-100), mean (SD)	30	64.6 (24.8)	40	77.8 (17.1)	.02[a]
RAND physical role limitations (0-100), mean (SD)	30	36.7 (43.9)	40	60.0 (42.7)	.03[a]
RAND emotional role limitations (0-100), mean (SD)	29	74.7 (41.5)	40	85.8 (33.7)	.24
RAND mental health (0-100), mean (SD)	30	72.1 (16.1)	40	78.7 (11.6)	.06
RAND vitality (0-100), mean (SD)	30	53.1 (22.9)	40	61.7 (15.4)	.08
RAND pain (0-100), mean (SD)	30	56.9 (25.5)	40	69.8 (19.2)	.03[a]
RAND general health perception (0-100), mean (SD)	29	46.0 (19.4)	40	54.0 (17.6)	.08
RAND health change (0-100), mean (SD)	30	43.3 (20.7)	40	52.5 (24.6)	.10
NRS[g] pain today (0-10), mean (SD)	29	4.3 (2.5)	40	2.5 (1.8)	.002[a]
NRS mean pain last 2 weeks (0-10), mean (SD)	30	4.4 (2.5)	40	3.1 (2.1)	.02[a]
NRS fatigue today (0-10), mean (SD)	30	4.8 (2.4)	40	3.0 (2.2)	<.001[a]
NRS mean fatigue last 2 weeks (0-10), mean (SD)	30	4.8 (2.4)	40	3.0 (2.2)	<.001[a]
MPCI-F[h] (4-32), mean (SD)	30	15.0 (4.8)	40	13.2 (3.0)	.08

[a] Significant differences (*P*<.05) between the group low and high users.

[b] PAM: Patient Activation Measurement.

[c] SMAS-S: short Self-Management Ability Scale.

[d] RASE: Rheumatoid Arthritis Self-Efficacy.

[e] PEPPI-5: Perceived Efficacy in Patient-Physician Interaction.

[f] RAND-36: General Health Status.

[g] NRS pain/fatigue: Numeric Rating scales pain/fatigue.

[h] Coping with fatigue: Modified Pain Coping Inventory for Fatigue.

Table 6. The estimated difference between the group with low and high usage of the intervention after sensitivity analysis at 6 months and 12 months after baseline.

Scales	6 months after baseline				12 months after baseline			
	T0-T1 change	95% CI	*P* value[a]	Cohen *d*	T0-T2 change	95% CI	*P* value[a]	Cohen *d*
PAM[b] (10-65)	2.4	–1.7 to 6.4	.24	0.12	0.0	–2.9 to 2.9	>.99	0.00
SMAS-S[c] (0-60)	–0.4	–3.4 to 2.7	.82	0.00	1.3	–2.0 to 4.5	.44	0.02
RASE[d] (28-140)	–1.7	–6.8 to 3.4	.52	–0.00	–0.6	–5.3 to 4.1	.81	0.00
PEPPI-5[e] (5-25)	–1.0	–2.5 to 0.5	.20	–0.11	–0.1	–1.9 to 1.7	.93	0.00
RAND[f] physical functioning (0-100)	9.2	–0.7 to 19.2	.07	0.01	2.2	–7.4 to 11.8	.65	0.00
RAND social functioning (0-100)	1.5	–8.4 to 11.4	.76	0.00	5.3	–6.7 to 17.4	.38	0.01
RAND physical role limitations (0-100)	7.4	–14.7 to 29.5	.51	0.00	3.7	–18.6 to 25.9	.74	0.00
RAND emotional role limitations (0-100)	16.1	–3.6 to 35.7	.11	0.01	–1.7	–24.5 to 21.0	.88	0.00
RAND mental health (0-100)	0.8	–6.3 to 7.9	.83	0.00	–4.2	–11.2 to 2.8	.24	–0.02
RAND vitality (0-100)	2.9	–5.6 to 11.5	.50	0.01	–1.2	–10.8 to 8.4	.81	0.00
RAND pain (0-100)	1.7	–4.5 to 12.9	.77	0.00	8.8	–3.0 to 20.6	.14	0.02
RAND general health perception (0-100)	2.9	–5.1 to 10.8	.48	0.01	9.7	0.8 to 18.5	.03[g]	0.02
RAND health change (0-100)	8.3	–4.0 to 20.5	.19	0.02	6.4	–7.9 to 20.6	.38	0.01
NRS[h] pain today (0-10)	0.0	–1.2 to 1.2	>.99	0.00	–0.6	–1.9 to 0.8	.41	–0.11
NRS mean pain last 2 weeks (0-10)	–0.6	–1.8 to 0.5	.29	–0.12	0.9	–2.3 to 0.6	.24	–0.16
NRS fatigue today (0-10)	0.2	–1.0 to 1.3	.73	0.03	–0.9	–2.2 to 0.5	.22	–0.14
NRS mean fatigue last 2 weeks (0-10)	0.2	–1.0 to 1.3	.78	0.03	–0.5	–1.7 to 0.8	.51	–0.08
MPCI-F[i] (4-32)	–0.2	–1.7 to 1.3	.76	–0.01	–0.4	–2.1 to 1.3	.67	–0.01

[a] Values represent outcomes of the ITT analysis without confounders. After adding confounders, no changes in values appear.

[b] PAM: Patient Activation Measurement.

[c] SMAS-S: short Self-Management Ability Scale.

[d] RASE: Rheumatoid Arthritis Self-Efficacy.

[e] PEPPI-5: Perceived Efficacy in Patient-Physician Interaction.

[f] RAND-36: General Health Status.

[g] Significant differences (*P*<.05) between control and intervention groups.

[h] NRS: numerical rating scale.

[i] MPCI-F: Modified Pain Coping Inventory for Fatigue.

Discussion

Principal Findings

This study aimed to evaluate the efficacy of a Web-based self-management enhancing program in patients with RA in an explorative trial on 6 outcomes: self-management behavior, self-efficacy, general health status, coping with fatigue, and the level of pain and fatigue. Results show no remarkable statistically significant difference between the intervention and control group. Moreover, effect sizes were low. Consequently, the results of this exploratory show no convincing trend regarding the efficacy of the program. This was unexpected as the theory-based intervention was carefully designed, according the IM steps, on the basis of patients support needs [37,38]. In addition, the range of outcome measures were selected carefully, and the study was well-performed. Randomization was successful, and the number of missing was limited. It was thought that the size was adequate for a pilot study (N=157).

Notably, the lack of a trend for a positive result is not in line with other studies, showing that self-management programs seem to be promising for patients with a chronic illness, including arthritis [15,35]. However, these studies cannot be compared with each other in a straightforward manner because of the various self-management approaches (eg, offering weekly vs nonweekly Web-based courses, with face-to-face help or without), various contents of the self-management programs, and the different outcome measures used in these studies [15,39]. For example, it is unexpected that our Web-based program yielded no results for RA patients, whereas the Web-based program evaluated by Lorig and colleagues [14] concluded that RA patients showed increased self-efficacy and improved health status for 4 of the 6 health status measures that were included [14]. These different results may be explained by the different questionnaires used for the same outcomes, that is, self-efficacy and health status.

Moreover, differences in the content and delivery of the programs could be a reason for the different results. Other programs focused on different topics (eg, pain/stress management, problem solving and nutrition, which were not covered by our program). In our program, patients received no help with logging into the program or using the program in contrast to the program described by Lorig et al [14], where patients received help and were encouraged to use the program. Patients could choose which modules to work through and follow it at their own speed. In the program described by Lorig et al [14], peer moderators helped patients log in and encouraged them to use the weekly program and moderate posts that patients could leave on the program website [14].

There are potentially 5 reasons for the lack of efficacy of our Web-based program: (1) the use of inappropriate outcome measures, (2) individual patients had no need for self-management support, (3) low usage of the program/high dropout of the intervention group, (4) inadequate embedding of the program in health care, and (5) not selecting the appropriate patients.

First, in the case of inappropriate outcome measures, it could be that the carefully selected validated questionnaires still did not exactly measure the pursued behavior changes formulated in the performance objectives. That is, the intervention aimed to result in specific self-management behaviors. The validated questionnaires comprised more generic questions and therefore did not exactly measure these specifically formulated behavioral changes in performance objectives (Table 1). However, it was expected that a positive significant result would be found on the RASE questionnaire, as this measures task-specific self-efficacy for patients with RA with items closely related to the specific formulated performance objectives. Finding no positive results suggests that it is possible that our intervention did not support patients in increasing their level of self-efficacy. This could mean that the absence of positive results is less driven by the choice of outcomes than by the other points discussed below.

Second, it could be that recruited patients did not have a perceived need for enhancing self-efficacy when they agreed to participate in the program. Although the program was developed on the basis of the support needs for self-management of RA patients, individual participating patients in this study were not asked whether, and if yes, what kind of support needs they had for self-management. It could be that patients differ in their needs and more tailoring toward individuals is needed, for example, preselection of the offered modules.

Third, the low usage of the program by patients in the intervention group could have resulted in finding only a significant effect on RAND-36 vitality, with a small effect size. The low usage of the program can have several reasons. As stated above, patients could have not felt a need for support. Another reason could be that patients were not motivated to change their behavior or had a negative attitude toward the Web-based program. The program comprised several elements to stimulate patients' usage of the program, such as persuasive texts or modeling videos. It could be that these elements did not work or that elements were lacking in the program. Moreover, the characteristics of the Web-based program, for example, attractiveness or the ease of logging in, are factors that could have influenced patient usage of the program. It was also notable that patients in the intervention group dropped out more than patients in the control group. A high dropout rate is a common finding in Web-based programs [40,41]. Crutzen et al (2015), gave as possible explanation for these higher dropout rates that patients in an intervention group have several expectations of the intervention. In cases where these expectations are not met or if patients feel the intervention is not supporting them, patients will refuse to fill in the measurements and will not revisit the program [42]. In this study, patients in the intervention group were significantly higher educated than in the control group. It could be that higher educated patients use more resources that could support them (eg, support of health professionals), which could lead to lower usage of the program.

Fourth, this program was not adequately embedded in patient care. Although nurses brought the program to the attention of intervention group patients during their consultation, they did not discuss the self-management topics of the program with patients to continue the support for self-management during consultations. It has been shown that self-management programs with the possibility of interacting with health professionals (blended care) can lead to positive results [14,43].

Fifth, it could be that there was a selection bias in this study. Rheumatologists selected patients with diagnosis RA, aged 18 years, or were invited by letter to participate in this study, in collaboration with rheumatologists until the required number of 190 patients were reached. Probably, rheumatologists mainly selected the patients who had a low functional disability (health assessment questionnaire) as in their opinion, these patients would benefit of a self-management program the most.

Recommendations

Given the results of this study, relevant recommendations for future studies and practice can be given. First, using a questionnaire with questions referring to the program objectives is recommended to measure the effects in patient behavioral change [38]. For example, one of the performance objectives of this Web-based program, *"set boundaries in their work situation,"* could be evaluated with an item like *"I'm able to set boundaries with my colleagues in my work situation"* (measuring skills). Patients can set their own objectives in the program, using goal setting as a strategy. Goal setting requires that patients set a clear, specific, and achievable goal to change their behavior. This concrete formulation of the goal ensures that the behavioral change is measurable [44].

Second, before inclusion, it is recommended to investigate whether patients have a need for self-management support and if so, what kind of support they need. A next step is to decide if patient support needs are handled in the program and to tailor the program to their support needs. This can avoid patients feeling that the program did not support them, which often results in no revisits. Investigation of support needs could take place over the telephone. This also offers the possibility of helping patients formulate their support needs, which is difficult to do in general.Third, to increase the usage of the program and limit dropout, during the development phase, it is important to pay attention to factors that could enhance usage of the program (first visit, staying on the website, revisits).Patients input, in combination with attention to dissemination, reach, adoption and implementation (emphasized in diffusion theory or RE-AIM theory), could be used to identify factors [45,46]. Moreover, qualitative research to explore the reasons for low usage should be conducted.

Fourth, to embed the program in regular health care, it is important that patient needs are also recognized by their rheumatologists or specialized rheumatology nurse and be used as a starting point during consultation. Nurses could also assist patients in performing exercises mentioned in the program, reminding patients to log on to the website and encourage patients to maintain their self-management behavior. Fifth, to increase the usage and efficacy of the program, a specific patient selection is needed. Further research is needed to assess which patient characteristics influence the use of a Web-based program and the outcomes, for example, by performing subgroup analysis among groups with a low or high functional disability or by assessing their level of motivation to use the program. This can determine which inclusion criteria should be used to select patients likely to benefit most.

Conclusions

In conclusion, although there is external evidence in favor of the efficacy of Web-based self-management interventions [14,15], it is not recommended to conduct a larger trial yet. As advised by the MRC framework, a detailed process evaluation of the program should be conducted to gain thorough insight into the implementation of the program, the working elements of the program and the usage of the program by patients, which could be both important conditions for the success of a self-management program. This could also satisfy the need for attention to the usage and the perceived impact of the program to find out for whom this sort of program will work [47].

Acknowledgments

The authors would like to thank the RA patients and Valesca Hulsman for her help with translating the RASE questionnaire.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 38MB - jmir_v21i4e12463_app1.pdf]

References

1. Hannan M. Epidemiologic perspectives on women and arthritis: an overview. Arthritis Care Res 1996 Dec;9(6):424-434. [doi: 10.1002/art.1790090603] [Medline: 9136285]
2. Sarzi-Puttini P, Salaffi F, di Franco M, Bazzichi L, Cassisi G, Casale R, et al. Pain in rheumatoid arthritis: a critical review. Reumatismo 2014 Jun 6;66(1):18-27 [FREE Full text] [doi: 10.4081/reumatismo.2014.760] [Medline: 24938192]
3. Repping-Wuts H, Uitterhoeve R, van Riel P, van Achterberg T. Fatigue as experienced by patients with rheumatoid arthritis (RA): a qualitative study. Int J Nurs Stud 2008 Jul;45(7):995-1002. [doi: 10.1016/j.ijnurstu.2007.06.007] [Medline: 17662291]
4. Repping-Wuts H, Fransen J, van Achterberg T, Bleijenberg G, van Riel P. Persistent severe fatigue in patients with rheumatoid arthritis. J Clin Nurs 2007 Nov;16(11C):377-383. [doi: 10.1111/j.1365-2702.2007.02082.x] [Medline: 17931330]
5. ter Wee MM, Lems WF, Usan H, Gulpen A, Boonen A. The effect of biological agents on work participation in rheumatoid arthritis patients: a systematic review. Ann Rheum Dis 2012 Feb;71(2):161-171. [doi: 10.1136/ard.2011.154583] [Medline: 21998122]
6. Connolly D, Fitzpatrick C, O'Toole L, Doran M, O'Shea F. Impact of fatigue in rheumatic diseases in the work environment: a qualitative study. Int J Environ Res Public Health 2015;12(11):a. [doi: 10.3390/ijerph121113807] [Medline: 26516896]
7. Wikstrom I, Isacsson A, Jacobsson LT. Leisure activities in rheumatoid arthritis: change after disease onsetassociated factors. Br J Occup Ther 2001;2(64):87-92. [doi: 10.1177/030802260106400206]
8. Giorgino KB, Blalock SJ, DeVellis RF, DeVellis BM, Keefe FJ, Jordan JM. Appraisal of and coping with arthritis-related problems in household activities, leisure activities, and pain management. Arthritis Care Res 1994 Mar;7(1):20-28. [doi: 10.1002/art.1790070106] [Medline: 7918722]
9. MacKinnon J, Avison W, McCain G. Pain and functional limitations in individuals with rheumatoid arthritis. Int J Rehabil Res 1994 Mar;17(1):49-59. [Medline: 7960328]
10. Nichol M, Zhang L. Depression and health-related quality of life in patients with rheumatoid arthritis. Expert Rev Pharmacoecon Outcomes Res 2005 Oct;5(5):645-653. [doi: 10.1586/14737167.5.5.645] [Medline: 19807590]

11. Radford S, Carr M, Hehir M, Davis B, Robertson L, Cockshott Z, et al. 'It's quite hard to grasp the enormity of it': perceived needs of people upon diagnosis of rheumatoid arthritis. Musculoskeletal Care 2008 Sep;6(3):155-167. [doi: 10.1002/msc.132] [Medline: 18649347]
12. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
13. Atkinson NL, Gol RS. The promise and challenge of eHealth interventions. Am J Health Beh 2002;26(6):494-503. [doi: 10.5993/AJHB.26.6.10] [Medline: 12437024]
14. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
15. Warsi A, Wang PS, LaValley MP, Avorn J, Solomon DH. Self-management education programs in chronic disease: a systematic review and methodological critique of the literature. Arch Intern Med 2004;164(15):1641-1649. [doi: 10.1001/archinte.164.15.1641] [Medline: 15302634]
16. Lorig K, González VM, Laurent D, Morgan L, Laris B. Arthritis self-management program variations: three studies. Arthritis Care Res 1998 Dec;11(6):448-454. [doi: 10.1002/art.1790110604] [Medline: 10030176]
17. Stinson J, McGrath PJ, Hodnett ED, Feldman BM, Duffy CM, Huber AM, et al. An internet-based self-management program with telephone support for adolescents with arthritis: a pilot randomized controlled trial. J Rheumatol 2010 Sep;37(9):1944-1952. [doi: 10.3899/jrheum.091327] [Medline: 20595280]
18. Or C, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
19. van der Vaart R, Drossaert CH, Taal E, van de Laar MA. Patient preferences for a hospital-based rheumatology Interactive Health Communication Application and factors associated with these preferences. Rheumatology (Oxford) 2011 Sep;50(9):1618-1626. [doi: 10.1093/rheumatology/ker161] [Medline: 21551221]
20. Zuidema R, van Gaal GB, van Dulmen S, Repping-Wuts H, Schoonhoven L. An online tailored self-management program for patients with rheumatoid arthritis: a developmental study. JMIR Res Protoc 2015;4(4):a. [doi: 10.1080/13548500410001721855] [Medline: 26706653]
21. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
22. Kok G, Schaalma H, Ruiter RA, van Empelen P, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 18824488]
24. Zuidema R, Repping-Wuts H, Evers AW, Van Gaal BG, Van Achterberg T. What do we know about rheumatoid arthritis patients' support needs for self-management? A scoping review. Int J Nurs Stud 2015 Oct;52(10):1617-1624. [doi: 10.1016/j.ijnurstu.2015.05.008] [Medline: 26117711]
25. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-T]
26. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
27. Uhlig T, Haavardsholm EA, Kvien TK. Comparison of the Health Assessment Questionnaire (HAQ) and the modified HAQ (MHAQ) in patients with rheumatoid arthritis. Rheumatology (Oxford) 2006 Apr;45(4):454-458. [doi: 10.1093/rheumatology/kei181] [Medline: 16287925]
28. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
29. Cramm J, Strating M, de Vreede VP, Steverink N, Nieboer A. Validation of the self-management ability scale (SMAS) and development and validation of a shorter scale (SMAS-S) among older patients shortly after hospitalisation. Health Qual Life Outcomes 2012;10:9. [doi: 10.1186/1477-7525-10-9]
30. Hewlett S, Cockshott Z, Kirwan J, Barrett J, Stamp J, Haslock I. Development and validation of a self-efficacy scale for use in British patients with rheumatoid arthritis (RASE). Rheumatology (Oxford) 2001;40(11):1221-1230. [doi: 10.1093/rheumatology/40.11.1221] [Medline: 11709605]
31. ten Klooster PM, Oostveen J, Zandbelt L, Taal E, Drossaert C, Harmsen E, et al. Further validation of the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns 2012 Apr;87(1):125-130. [doi: 10.1016/j.pec.2011.07.017] [Medline: 21889864]
32. Vander Zee KI, Sanderman R, Heyink JW, de Haes H. Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. Int J Behav Med 1996;3(2):104-122. [doi: 10.1207/s15327558ijbm0302_2] [Medline: 16250758]

33. Kraaimaat FW, Evers AW. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med 2003;10(4):343-363. [doi: 10.1207/S15327558ijbm1004_5] [Medline: 14734263]
34. Arain M, Campbell M, Cooper C, Lancaster G. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010 Jul 16;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
35. Lorig K, Sobel D, Ritter P, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
36. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis For The Behavioral Sciences Third Edition. New Jersey: Routledge; 2019.
37. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
38. Bartholomew L. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco: Jossey-Bass; 2006.
39. Trappenburg J, Jonkman N, Jaarsma T, van Os-Medendorp H, Kort H, de Wit N, et al. Self-management: one size does not fit all. Patient Educ Couns 2013;92(1):134-137. [doi: 10.1016/j.pec.2013.02.009]
40. Crutzen R, Viechtbauer W, Spigt W, Kotz D. Differential attrition in health behaviour change trials: a systematic review and meta-analysis. Psychol Health 2015 Jan;30(1):122-134. [doi: 10.1080/08870446.2014.953526] [Medline: 25109224]
41. Brouwer W, Kroeze W, Crutzen R, de Nooijer N, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 6;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
42. Crutzen R, Viechtbauer W, Kotz D, Spigt M. No differential attrition was found in randomized controlled trials published in general medical journals: a meta-analysis. J Clin Epidemiol 2013 Sep;66(9):948-954. [doi: 10.1016/j.jclinepi.2013.03.019] [Medline: 23790724]
43. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
44. Hurn J, Kneebone I, Cropley M. Goal setting as an outcome measure: a systematic review. Clin Rehabil 2006 Sep;20(9):756-772. [doi: 10.1177/0269215506070793] [Medline: 17005500]
45. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/AJPH.89.9.1322] [Medline: 10474547]
46. Rogers EM. Diffusion of Innovations. New York: Simon & Schuster; 2003.
47. Campbell N, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. Br Med J 2007 Mar 3;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]

Abbreviations

IM: intervention mapping
M-HAQ: Modified Health Assessment Questionnaire
MPCI-F: Modified Pain Coping Inventory for Fatigue
MRC: Medical Research Council
NRS: Numeric Rating Scale
PAM-13: Patient Activation Measurement
RA: rheumatoid arthritis
RASE: Rheumatoid Arthritis task-specific Self-Efficacy Questionnaire
RCT: randomized controlled trial
SMAS-S: Short Self-Management Ability Scale

Edited by G Eysenbach; submitted 12.10.18; peer-reviewed by P van Riel, C Varsi; comments to author 11.11.18; revised version received 29.11.18; accepted 29.11.18; published 30.04.19.

Please cite as:
Zuidema R, van Dulmen S, Nijhuis-van der Sanden M, Meek I, van den Ende C, Fransen J, van Gaal B
Efficacy of a Web-Based Self-Management Enhancing Program for Patients with Rheumatoid Arthritis: Explorative Randomized Controlled Trial
J Med Internet Res 2019;21(4):e12463
URL: https://www.jmir.org/2019/4/e12463/
doi:10.2196/12463
PMID:31038461

Original Paper

Mood Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study

Chul-Hyun Cho[1*], MD, PhD; Taek Lee[2*], PhD; Min-Gwan Kim[3], MS; Hoh Peter In[3], PhD; Leen Kim[1], MD, PhD; Heon-Jeong Lee[1], MD, PhD

[1]Korea University College of Medicine, Department of Psychiatry, Seoul, Republic of Korea
[2]Sungshin University, Department of Convergence Security Engineering, Seoul, Republic of Korea
[3]Korea University College of Informatics, Department of Computer Science and Engineering, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Heon-Jeong Lee, MD, PhD
Korea University College of Medicine
Department of Psychiatry
Anam Hospital, Korea University College of Medicine
Seoul,
Republic of Korea
Phone: 82 029205815
Email: leehjeong@korea.ac.kr

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2019/10/e15966

Abstract

Background: Virtually, all organisms on Earth have their own circadian rhythm, and humans are no exception. Circadian rhythms are associated with various human states, especially mood disorders, and disturbance of the circadian rhythm is known to be very closely related. Attempts have also been made to derive clinical implications associated with mood disorders using the vast amounts of digital log that is acquired by digital technologies develop and using computational analysis techniques.

Objective: This study was conducted to evaluate the mood state or episode, activity, sleep, light exposure, and heart rate during a period of about 2 years by acquiring various digital log data through wearable devices and smartphone apps as well as conventional clinical assessments. We investigated a mood prediction algorithm developed with machine learning using passive data phenotypes based on circadian rhythms.

Methods: We performed a prospective observational cohort study on 55 patients with mood disorders (major depressive disorder [MDD] and bipolar disorder type 1 [BD I] and 2 [BD II]) for 2 years. A smartphone app for self-recording daily mood scores and detecting light exposure (using the installed sensor) were provided. From daily worn activity trackers, digital log data of activity, sleep, and heart rate were collected. Passive digital phenotypes were processed into 130 features based on circadian rhythms, and a mood prediction algorithm was developed by random forest.

Results: The mood state prediction accuracies for the next 3 days in all patients, MDD patients, BD I patients, and BD II patients were 65%, 65%, 64%, and 65% with 0.7, 0.69, 0.67, and 0.67 area under the curve (AUC) values, respectively. The accuracies of all patients for no episode (NE), depressive episode (DE), manic episode (ME), and hypomanic episode (HME) were 85.3%, 87%, 94%, and 91.2% with 0.87, 0.87, 0.958, and 0.912 AUC values, respectively. The prediction accuracy in BD II patients was distinctively balanced as high showing 82.6%, 74.4%, and 87.5% of accuracy (with generally good sensitivity and specificity) with 0.919, 0.868, and 0.949 AUC values for NE, DE, and HME, respectively.

Conclusions: On the basis of the theoretical basis of chronobiology, this study proposed a good model for future research by developing a mood prediction algorithm using machine learning by processing and reclassifying digital log data. In addition to academic value, it is expected that this study will be of practical help to improve the prognosis of patients with mood disorders by making it possible to apply actual clinical application owing to the rapid expansion of digital technology.

(J Med Internet Res 2019;21(4):e11029) doi:10.2196/11029

KEYWORDS

mood disorder; circadian rhythm; projections and predictions; machine learning; digital phenotype; wearable device

Introduction

Background

Mood disorders, such as major depressive disorder (MDD) and bipolar disorder (BD), are common, with recurrent mood episodes and substantial chronicity. Patients with mood disorders suffer from a high disease burden, disrupted functional levels, and increased suicide risk [1,2]. It is crucial to use a coordinated approach to enhance prognosis by proactively managing symptoms and preventing recurrences. For successful prognosis enhancement, a new therapeutic approach is needed to assess, analyze, and manage the patient's daily condition, in addition to conventional pharmacotherapeutic and psychotherapeutic approaches.

The circadian rhythm mechanism has been identified as an important factor in the onset and aggravation of mood disorders [3-5]. It has been reported that disturbances in circadian rhythms could be a unique clinical manifestation of mood disorders, and phase shift of circadian rhythms can serve as a marker for mood disorders [6,7]. Digital technology and machine learning have recently shown remarkable progress, bringing substantial changes into the lives of individuals [8]. In medicine, the use of digital phenotypes obtained with wearable technology or mobile devices has been reported [8-10]. Within psychiatry, research based on digital technology and machine learning has recently been introduced. This new research methodology is expected to overcome many limitations of existing psychiatric research through the precise analysis of clinical information obtained from various digital phenotypes. Digital phenotyping allows us to more closely and continuously measure information on a variety of biometrics, such as mood, activity, heart rate, and sleep, in the patient's daily life and to connect these with clinical symptoms.

Objectives

Using chronobiological concepts of mood disorders, in this prospective study, we collected long-term clinical mood logs and passively collected data on activity, sleep, light exposure, and heart rate in patients with mood disorders. Data were collected continuously through wearable devices and mobile technologies. We then analyzed the data to determine whether mood states or episodes could be predicted using only the automatically recorded data, without any knowledge on mood information, by machine learning.

Methods

Recruitment and Study Design

From March 2015 to December 2017, 55 patients (27 females and 28 males) diagnosed with a major mood disorder (MDD=18 subjects, BD I=18, and BD II=19) according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), [11] were recruited from the Korea University Anam Hospital as part of the Mood Disorder Cohort Research Consortium (MDCRC) study (ClinicalTrials.gov: NCT03088657). Each patient had different days of participation period during the entire study period. The MDCRC study is a multicenter prospective observational cohort study investigating early-onset mood disorders in Korea, and its design and protocol have been reported previously [12]. The average age (SD) of the patients, age at first onset of mood disorder, and age at first psychiatric treatment was 25.92 (SD 4.78), 17.87 (SD 4.80), and 20.69 (SD 4.13) years, respectively (Multimedia Appendix 1). The study was approved by the Institutional Review Board of Korea University Anam Hospital and conducted in accordance with the Declaration of Helsinki. All participants provided informed written consent before enrollment after receiving a full explanation of the study.

Assessment

In addition to standard clinical assessments conducted at regular intervals, including demographic and clinical data from investigators and patients, we provided an eMoodchart smartphone app developed on our own and a wearable activity tracker (Fitbit Charge HR or 2, Fitbit Inc). The smartphone app had an alert set for 9:00 o'clock every night, when patients recorded a simple, intuitive assessment of their daily mood state (−3 to +3) on the eMoodchart app. At every clinical assessment, a clinician reviewed the eMoodchart and determined the mood episodes that had occurred since the previous clinic visit. The between-visit mood episode evaluation was based on the eMoodchart and the patient interview according to DSM-5 [12]. For the comparison of basic features according to mood states, daily mood scores (−3 to +3) were converted to the absolute mood score (AMS; 0 to 3). When using the original recorded mood score as it is for analysis, it was difficult to reflect all kinds of mood states such as elated, depressed, or mixed, and it could add complexity to the overall trends. For the overall trend analysis, we decided to simply rearrange the mood to be stable or unstable in 2 directions. In other words, the higher the AMS, the mood can be regarded as worse and unstable (more depressed or more manic), and the lower the AMS, the mood can be regarded as more stable. For smartphones using the Android operating system (40 patients), the app could also detect light exposure, using a built-in sensor. The activity trackers, worn continuously, collected passive data related to activity, sleep, and heart rate, which were then obtained by the researchers from the Fitbit cloud server. The practicality and validity of using the Fitbit series for clinical research [13] and clinical results have been reported elsewhere [14]. As the smartphones are easily and frequently used in everyday life, there have been previous studies using the built-in sensors in smartphones [15-17].

Datasets

During data collection, missing data were occasionally encountered for a variety of reasons (failure to complete the eMoodchart and removal of the activity tracker continuously over 24 hours). During the data collection period, we could originally collect 17,542 sample days from 55 patients, but the total number of 2003 days remained after removing the days with any single missing variable. In our analysis, we used only the complete dataset of 2003 days. Our dataset for prediction modeling has 130 variables (features) plus a class label of the mood state. We excluded the entire row if the row had a column with any missing value among the 131 columns. Among all the missing value counts, 16.8% was about light-related features, 9.1% was about step-related features, 43.9% was about sleep-related features, 29.6% was about heart rate–related features, and 0.3% was about the mood score record; perhaps, many people did not want to wear the Fitbit during sleep time. Heart rate–related features need the past 48-hour sequential data to be computed so that those features are not resistant against some missing data.

It has been reported that mood state can be affected by the disruption of circadian rhythms [6,7]; therefore, we focused on identifying a set of features that would capture such disruptions. To achieve this, we focused on basic features derived from the 4 main data collection categories: (1) light exposure, (2) steps, (3) sleep, and (4) heart rate (Multimedia Appendix 2).

To calculate light exposure, we took the average value of light exposure levels observed during 2 timeslots of interest: bedtime and daytime. As the lengths of day and night change according to the seasons, *bedtime* was defined as the period from 8 hours before sunrise until sunrise the next day and *daytime* was defined as the period from sunrise to sunset each day. It is not easy to adequately reflect seasonal changes, but because light exposure in the early morning is the most important time-giver (zeitgeber) in the daily circadian rhythm, sunrise time is a key criterion for defining *daytime* and *bedtime*. Therefore, according to the seasonal changes, *bedtime* was set at a constant total sleep time but variable at the time of sleep onset and wake up and *daytime* was set at total times when a person could be exposed to sunlight. To measure activity levels, we collected step data that was calculated as total steps within the bedtime and daytime timeslots. Sleep data such as sleep length and quality and sleep onset and offset were also obtained from the Fitbit report. With regard to heart rate, variations in heart rate follow a circadian rhythm, with elevated rates during the daytime and lower rates at night. A cosinor analysis (cosine curve fitting) was performed on 48 consecutive hours of heart rate data, and 4 representative parameters were generated: amplitude, acrophase (*peak*), mesor (*mean*), and r-squared value (*strength*).

Finally, we extracted extended features from the 4 basic categories, which integrated data across multiple days. In constructing the prediction model, these features are used as predictors for mood state or episode. To predict mood in the near future, it can be helpful to look at snapshots of previous days; perhaps, people could be affected by mood changes in these preceding days. Therefore, we extended the daily snapshot feature to simultaneously include the previous consecutive 3, 6, and 12 days. For example, if today's date is d, then the mean value of the past 3 days would be from d-2 to d. In this way, the SD (*stdev*) and gradient coefficient (ie, a parameter gained from linear regression, *gradient*) can be computed for the extended features. The names of all features in the across-period perspective had a suffix including one of the 4 element names. The suffix terms also included the 3 elements describing the statistical perspective for the given period: *mean*, *stdev*, and *gradient*. Ultimately, we acquired 130 features (=13 basic features for every day+[13 basic features×3 types of the past periods×3 types of statistics for those periods]) for the data collected each day.

Development and Verification of the Mood State and Episode Prediction Algorithms

To train the mood prediction model, we used a supervised learning algorithm, random forest [18], that operates by constructing a multitude of decision trees at training time and outputting a class that is the mode of the classes of the individual trees. The random forest algorithm requires a training dataset that consists of a feature vector set, $X=x_1, \ldots, x_n$, with a corresponding class set, $Y=y_1, \ldots, y_n$, where n is the number of training data samples (ie, n=2003 in our study). The feature vector x_i has the form $f_1, f_2, \ldots, f_m$, where m is 130 and f_i has a feature value of the circadian rhythm. The class variable y_i has one of 2 mood states: *biased mood state* or *neutral mood state* for mood state prediction within the next 3 days and has one of 4 episodes: *depressive episode (DE)*, *manic episode (ME)*, *hypomanic episode (HME)*, or *no episode (NE; same as euthymic period)* for mood episode prediction. The mood state was defined as *neutral mood state* if the average AMS for the following 3 days is within the bottom 50% (low AMS) of all the observed AMS. Conversely, the mood state was defined as *biased mood state* if the average AMS for the following 3 days is within the top 50% (high AMS). The mood episode was determined in the *between-visit mood episode evaluation* conducted by the clinician [12]. Patients in the study experienced 57 DEs (major: 46, minor: 3, and brief: 8), 11 MEs, and 13 HMEs.

Performance of the trained prediction model was evaluated by assessing the model's accuracy, sensitivity, specificity, and the area under the curve (AUC) [19]. In a machine learning evaluation process, some portion of data is used for model training and the other portion is used for the model test. Training data should not include future measurements relative to the test data. To take into account such a temporal nature of the data and get a reliable evaluation statistic, we designed the model evaluation process as follows: first, data were sorted over the timeline. For an arbitrary time t on the timeline, a prediction model was trained using data on days d[t-p, t] and tested using data on days d[t+1, t+q], where p and q are the time period of days for model training and for model test, respectively. It is possible that the model performance can be changed depending on a different size of p or q. Therefore, to find a proper setting, we repeated and monitored experiments with changing p from 3 to 300 days and changing q from 3 to 30 days. Consequently, *P*=18 and q=3 were found as the best combination in our experiment setting (Multimedia Appendix 3), which implies that a short period such as 3 days is the most reasonable and effective setting in terms of predicting any distant future mood in our experiment. Therefore, for the performance evaluation of our proposed prediction model, we used the found parameter setting of p and q throughout the paper. Second, to get a reliable evaluation result, we needed to repeat enough evaluation rounds (ie, a round of model training and model test) so that we repeatedly measured performance metrics by moving t from the beginning to the end of the data over the timeline with the found parameter setting. Thus, the reported figures of sensitivity, specificity, accuracy, and AUC in the paper are average statistics from the repeated evaluation rounds. Finally, for a comparison between a general model and personalized model, a general model was developed using other people of the whole data and a personalized model was developed using only individual data. For data processing and model evaluation, we used a Python library tool, scikit-learn [20].

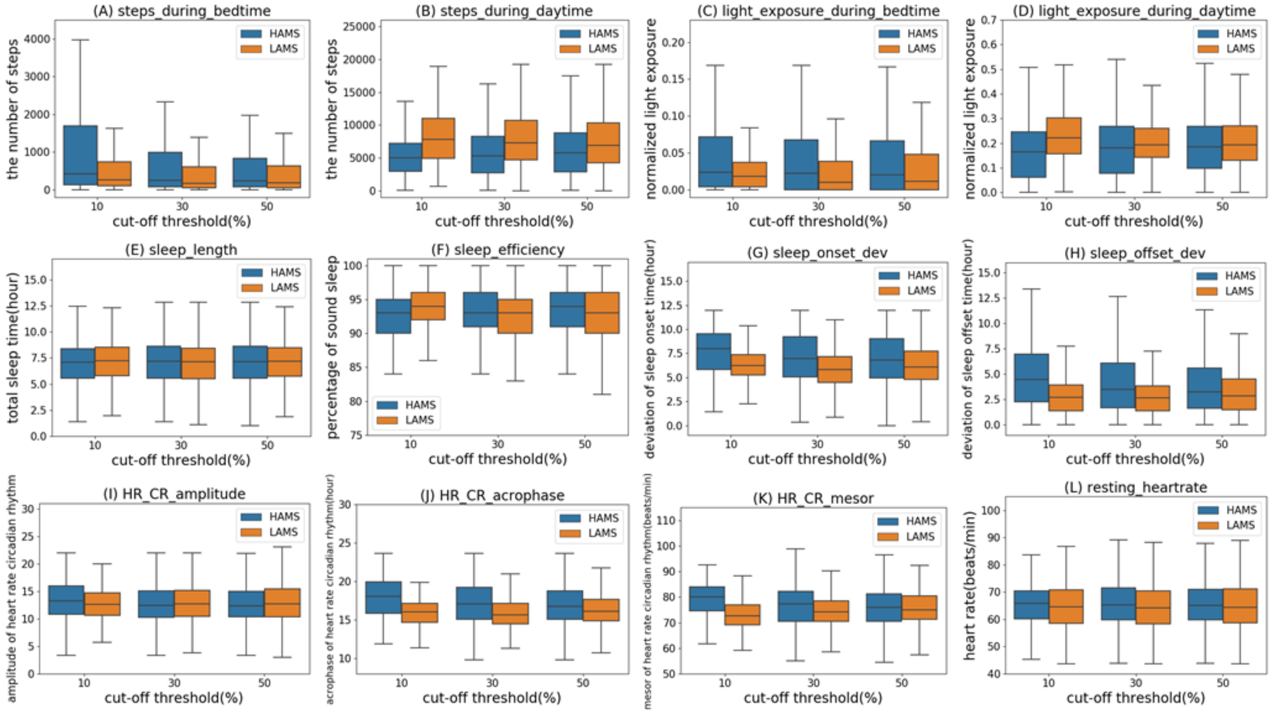
Results

Comparison of Main Basic Features According to Mood State

To confirm the appropriateness of the variables processed from a circadian rhythm perspective, we performed an exploratory review of the basic features by comparing them according to mood state. By sorting the collected data according to the AMS (0~3), mood state categories were created according to the AMS distribution and features belonging to them were compared. High and low AMS (HAMS or LAMS, respectively) days were grouped into the upper or lower 10%, 30%, and 50% thresholds of the distribution Therefore, the sum of the upper 50% of HAMS and the lower 50% of LAMS becomes the whole distribution of the AMS for each day of data. As presented in Figure 1, for each corresponding pair of threshold groups, we compared basic digital phenotypic features.

Activity and light exposure during bedtime showed a higher tendency in the HAMS groups than the LAMS groups. Conversely, activity and light exposure during daytime showed a higher tendency in LAMS (Figure 1). Interestingly, total sleep time and sleep quality did not show meaningful differences between the groups (Figure 1), although the regularity of sleep onset and offset times were disrupted in the HAMS groups (Figure 1), indicating that the regularity of the sleep-wake cycle is closely related to mood state. When the heart rate circadian rhythm was analyzed, it was found that the acrophase showed a remarkable difference between the HAMS and LAMS groups (Figure 1), suggesting that a misaligned or shifted heart rate acrophase could be a useful feature for determining mood state.

Figure 1. Comparison of basic feature distributions between high and low absolute mood score (HAMS and LAMS) groups. HAMS and LAMS were grouped into the upper or lower 10%, 30%, and 50% thresholds of the distribution. For each corresponding threshold group, we compared the main basic features of activity, light exposure, sleep, and heart rate (HR) related to the circadian rhythm (CR). The number in the parentheses on the horizontal axis means a cut-off threshold to distinguish HAMS and LAMS. (A) Steps_during_bedtime and (B) steps_during_daytime refer to activity levels of subjects during bedtime or daytime in each threshold HAMS or LAMS group. (C) Light_exposure_during_bedtime and (D) light_exposure_during_daytime refer to relative level of light exposure of subjects during bedtime or daytime. (E) Sleep_length and (F) sleep_efficiency refer to total sleep time (hours) and sleep efficiency (%), and (G) sleep_onset_dev and (H) sleep_offset_dev refer to regularity of sleep onset and offset time. (I) HR_CR_amplitude, (J) HR_CR_acrophase, and (K) HR_CR_mesor refer to the value of amplitude, acrophase, and mesor of cosine curve fitted HR, respectively. (L) resting_heart rate refers to the lowest HR at resting state during each day in the samples.



Performance Evaluation of the Mood State Prediction Model (Neutral or Biased Mood State)

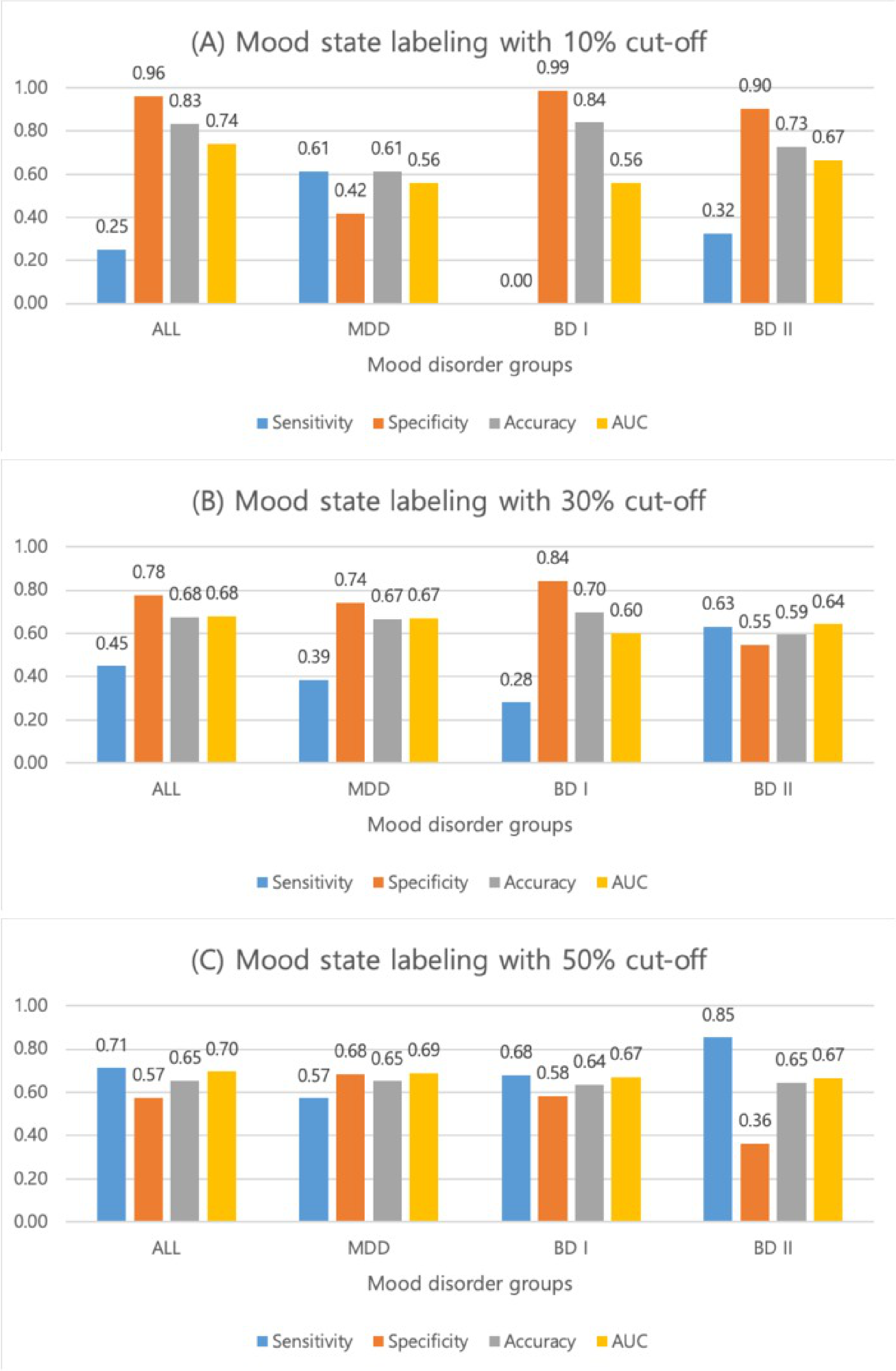
In model construction, we used 2-class labels for distinguishing a mood state, which were determined by a 10%, 30%, and 50% cut-off threshold (eg, 10% of the highest AMS is labeled as *biased mood state* and the rest of the 90% AMS is labeled as *neutral mood state* in the 10% cut-off case). We tested the model performance in the 3 different conditions of thresholds. As seen in Figure 2, we confirmed that the mood state prediction model performed better than a random prediction model, as the AUC values for the 3 patient groups, as well as for all groups combined, were all higher than 0.5. In the case of mood state labeling with a 50% cut-off (Figure 2), the prediction accuracy for all patients and for patients with MDD, BD I, or BD II was 65%, 65%, 64%, and 65%, respectively. Sensitivity was 71%, 57%, 68%, and 85%, specificity was 57%, 68%, 58%, and 36%, and AUC values were 0.7, 0.69, 0.67, and 0.67 for predicting mood states in all patients and in MDD, BD I, and BD II patients, respectively. Note that the ROC curves of Figure 2 presented in Multimedia Appendix 4 and additional information about the variance of the model performance in each evaluation round are reported in Multimedia Appendix 5. The number of samples used in the model construction for each case is reported in Multimedia Appendix 6.

To understand the quality of predictions using partially observed covariates of features, we evaluated each performance of the partial model construction with the whole patient data, supposing that some missing features were removed. The AUC performance of each partially constructed model was 0.684 without steps-related features, 0.687 without sleep-related features, 0.683 without heart rate–related features, and 0.683 without light-related features (more details present in Multimedia Appendix 8). In the impact analysis of missing features, heart rate– and light-related features were of a highly negative impact in terms of performance reduction.

To investigate the contribution of various features to the mood state predictions, we sorted the importance of influential features for prediction, depending on the patient group, as shown in Figure 3. The higher value in importance, the more frequently the feature is selected in a decision tree construction. To compute the feature importance, we used a Python library, scikit-learn [20], and referred to the code. In Figures 3 and 4, the color coding means the direction of feature effect was measured with Pearson correlation coefficients and the color-magnitude means relative strength of the correlation. Red color means a positive correlation with AMS (ie, the higher the feature value, the mood state tends to be more unstable.). Conversely, blue color means a negative correlation with AMS (ie, the higher the feature value, the less unstable the mood.). Gray color means the absolute coefficient values are less than 0.1, so it is hard to say any direction of the effect. Figures 3 and 4 have error variances of 1 SD with a solid black line at the end of each bar.

In the whole subject group and BD I group, the average circadian rhythm of heart rate (HR_CR_mesor) and deviation of sleep onset time (sleep_onset_dev) were the top influential features (Figure 3) and steps during bedtime were the most influential in the MDD group (Figure 3). Heart rate amplitude was the most influential in the BD II group (Figure 3).

Figure 2. The performance evaluation of the mood state prediction model. The mood state prediction model outputs one of 2 mood states (ie, biased mood state or neutral mood state) and whether the model outcomes that were correctly matched with the ground truth (ie, the known actual mood states) was tested. The mood performance was evaluated in terms of the 4 performance evaluation metrics: sensitivity, specificity, accuracy, and area under the curve with the 3 different ground truth labeling criterion: 10%, 30%, and 50% cut-offs in absolute mood score distribution. (A) The performance evaluation result in the case of mood state labeling with 10% cut-off, (B) the performance evaluation result in the case of mood state labeling with 30% cut-off, and (C) the performance evaluation result in the case of mood state labeling with 50% cut-off. MDD: major depressive disorder; BD I: bipolar I disorder; BD II: bipolar II disorder; AUC: area under the curve.



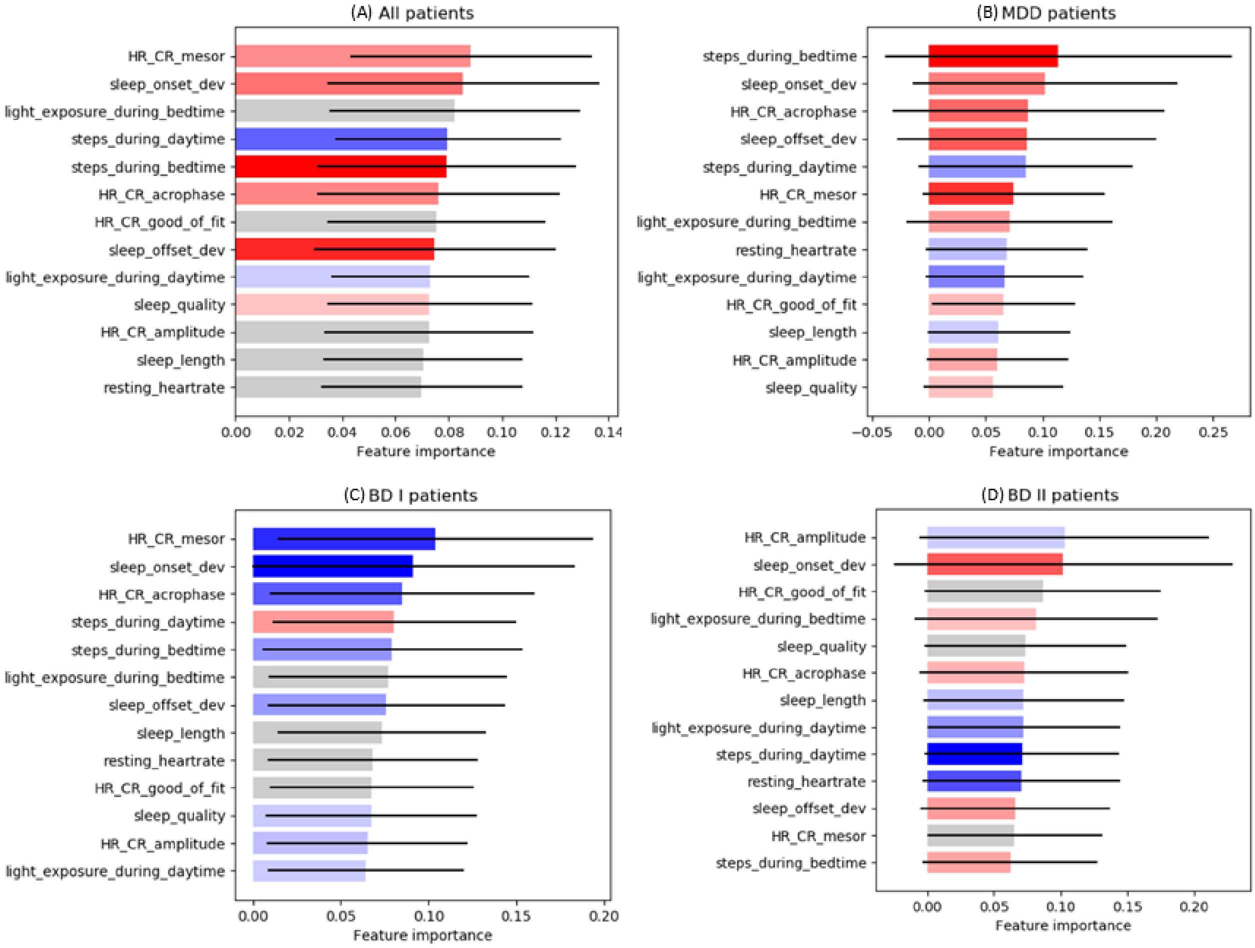
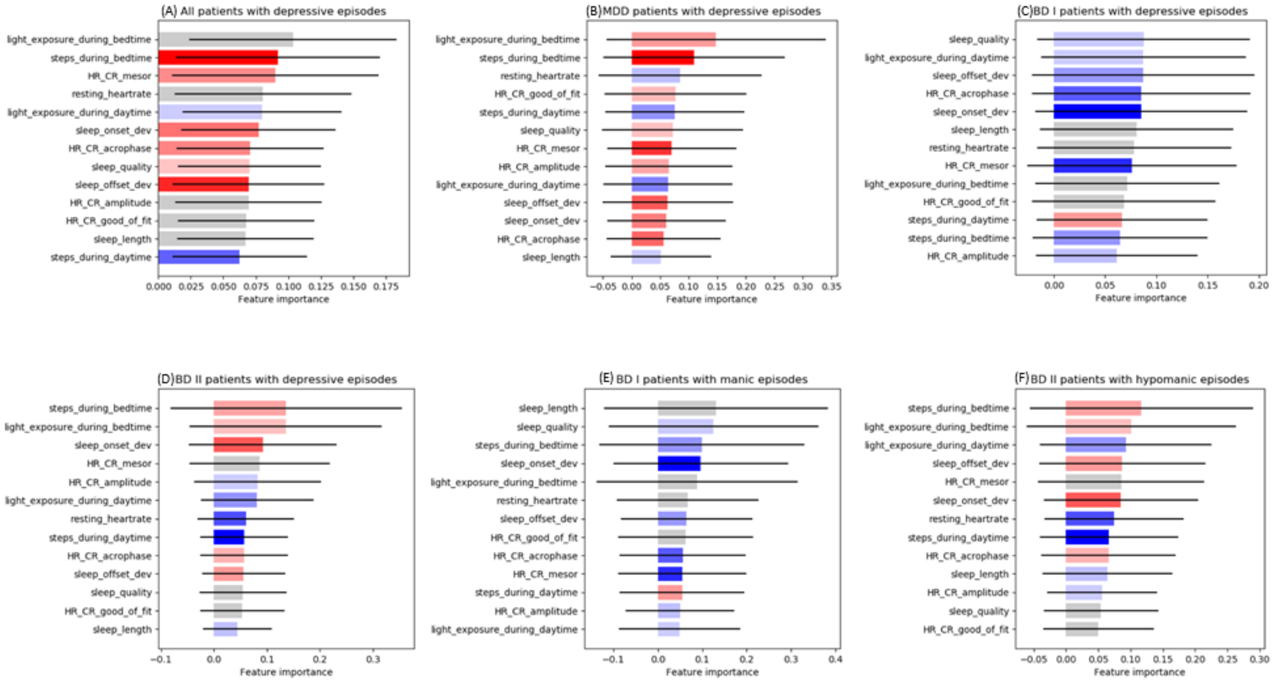


Figure 3. Representative influential features of the mood state prediction model. The mood state prediction model uses several categories of features and different features have different contribution to prediction performance. Each of the feature categories was compared in terms of its importance (contribution perspective). Each bar in the graph means relative importance of a feature category compared to the other bars. The black sold lines at the end of each bar means one standard deviation error range, and the color coding of each bar means the direction of feature effect; therefore, sum of all the bar lengths is one (100%). (A) The comparison analysis was conducted with data from all patients. (B) The analysis was conducted only with data from major depressive disorder patients. (C) The analysis was conducted only with data from bipolar I disorder patients. (D) The analysis was conducted only with data from bipolar II disorder patients. MDD: major depressive disorder; BD I: bipolar I disorder; BD II: bipolar II disorder; HR: heart rate; CR: circadian rhythm.

Figure 4. Representative influential features of the mood episode prediction model. The mood episode prediction model uses several categories of features, and different features have different contributions to prediction performance. Each of the feature categories was compared in terms of its importance (contribution perspective). Each bar in the graph means relative importance of a feature category compared with the other bars. The black sold lines at the end of each bar means 1 SD error range, and the color coding of each bar means the direction of feature effect; therefore, the sum of all the bar lengths is one (100%). (A) The comparison analysis was conducted with data from all patients with depressive episodes. (B) The analysis was conducted only with data from patients with major depressive disorder with depressive episodes. (C) The analysis was conducted only with data from patients with bipolar type I disorder with depressive episodes. (D) The analysis was conducted only with data from patients with bipolar type II disorder with depressive episodes. (E) The analysis was conducted only with data from patients with bipolar type I disorder with manic episodes. (F) The analysis was conducted only with data from patients with bipolar type II disorder with hypomanic episodes. MDD: major depressive disorder; BD I: bipolar type I disorder; BD II: bipolar type II disorder; HR: heart rate; CR: circadian rhythm.



Performance Evaluation of the Mood Episode Prediction Model

As seen in Table 1, the mood episode prediction model was determined to perform better than a random prediction model according to AUC values. Basically, AUC cannot be calculated for multiclassification but for binary classification. Therefore, we merged all the samples not in the target class but into the remaining class for multiclasses of mood episodes. The average accuracies in all patients and in those with MDD, BD I, and BD II were 89.3%, 73.1%, 85.1%, and 78.5%, respectively. For all patients, prediction accuracy for NE, DE, ME, and HME was 85.3%, 87%, 94%, and 91.2%, respectively. Sensitivity was 93%, 48%, 25.2%, and 30.6%, specificity was 59.9%, 95.6%, 99.6%, and 99.6%, and the AUC value was 0.87, 0.87, 0.958, and 0.912 for predicting NE, DE, ME, and HME, respectively. For patients with MDD, the prediction accuracy was 75.1% and 71.2%, sensitivity was 93.5% and 40.9%, specificity was 39.5% and 87.8%, and the AUC value was 0.781 and 0.798 for NE and DE, respectively. For patients with BD I, prediction accuracy was 84%, 83.1%, and 88.3%, sensitivity was 95.4%, 24.6%, and 20.7%, specificity was 39.3%, 97%, and 99.2%, and the AUC value was 0.84, 0.839, and 0.933 for NE, DE, and ME, respectively. For patients with BD II, prediction accuracy was 82.6%, 74.4%, and 87.5%, sensitivity was 84.5%, 64.1%, and 66.9%, specificity was 76.2%, 72.2%, and 98.4%, and the AUC value was 0.919, 0.868, and 0.949 for NE, DE, and HME, respectively. The number of samples used in the model construction for each case is reported in Multimedia Appendix 6. Figure 4 provides information on the important influential features in prediction performance capability in the diagnostic groups.

Table 1. The performance evaluation of the mood episode prediction model.

Episodes and measures	All patients	Major depressive disorder	Bipolar type I disorder	Bipolar type II disorder
No episodes (euthymic period)				
Samples, %	82	92.5	85	66.9
Sensitivity	0.93	0.965	0.954	0.845
Specificity	0.599	0.395	0.393	0.762
Accuracy	0.853	0.751	0.84	0.826
AUC[a]	0.87	0.781	0.84	0.919
Depressive episodes				
Samples, %	12.3	7.4	11.8	18.2
Sensitivity	0.48	0.409	0.246	0.641
Specificity	0.956	0.878	0.97	0.722
Accuracy	0.87	0.712	0.831	0.744
AUC	0.87	0.798	0.839	0.868
Manic episodes				
Samples, %	1.1	—[b]	3	—
Sensitivity	0.252	—	0.207	—
Specificity	0.996	—	0.992	—
Accuracy	0.94	—	0.883	—
AUC	0.958	—	0.933	—
Hypomanic episodes				
Samples, %	4.4	—	—	14.8
Sensitivity	0.306	—	—	0.669
Specificity	0.996	—	—	0.984
Accuracy	0.912	—	—	0.875
AUC	0.912	—	—	0.949
Average accuracy	0.893	0.731	0.851	0.785

[a] AUC: area under the curve.

[b] Not applicable.

General Model Versus Personalized Model

When constructing a machine learning model, the model is supposed to feed as much data as possible for the purpose of general knowledge learning. If a prediction model is learned by much data of the other people, it is called a general model in this paper. Conversely, if a prediction model is learned by personal data, it is called a personalized model. A general model can have comprehensive knowledge, but it is not specialized to an individual. A personalized model has specific knowledge especially for an individual. It is specialized to one so it is not applicable to the other one. It is a possible idea that a personalized model can improve the prediction accuracy of personal mood change. Therefore, as mood and related features showed various manifestations for each subject, we compared the performance of the personalized and general models in terms of accuracy for 4 prediction scenarios: (1) mood state for the next 3 days (*future mood state*), (2) DE, (3) ME, and (4) HME (Multimedia Appendix 7). For the future mood state, the personalized model outperformed the general model in 100% of the comparisons. The average degree to which the personalized model outperformed was 23.8%, meaning that all the cases of the personalized model predictions were 23.8% more accurate, on average, than those of the general model. In the scenario cases of episode prediction, the personalized models almost perfectly outperformed the general model as well.

Discussion

Principal Findings

In an exploratory review, we found several basic features from the passively collected data that showed a clear difference between the HAMS and LAMS groups. Light exposure is important because light is a central modulator of sleep, mood, and circadian rhythms [21]. Activity at appropriate times of the day and appropriate amounts could be one kind of social zeitgeber (*time-giver*) for the maintenance of circadian rhythms and mood [22]. Activity and light exposure, which are basic and important in terms of human circadian rhythms, consistently show clear and distinct differences according to the mood state. This suggests that it would be useful for patients with mood disorders to manage their activity levels and exposure to light to coordinate with their circadian rhythm to maintain a stable mood state. It is clinically significant that irregularities in the sleep-wake cycle are common in abnormal mood states [23]. The group differences in the heart rate acrophase identified in this study can be considered in the same context. Confirming the findings in Figure 1, we demonstrated that basic features related to circadian rhythms can meaningfully reflect the mood state.

The overall prediction accuracy for the mood state was relatively good. Interestingly, the sensitivity in BD II was markedly higher and the specificity lower than in other groups. This result may reflect characteristic features of BD II such as a close relationship to circadian rhythm disturbances as well as very common and sensitive mood changes in BD II. In the mood episode prediction model, overall prediction accuracy was quite good across all patient groups. The sensitivity to predict a DE was low across all groups except BD II (ie, all patients, MDD patients, and BD I patients). This may be because a DE is likely to be influenced by diverse factors including disturbance of circadian rhythms, socioeconomic stress, and interpersonal problems [6,24]. The prediction performance was markedly better in the BD II group for all episodes. In other words, the HME and DE in patients with BD II showed clearly distinguishable features compared with the euthymic period. BD II has been proposed as a distinctive major mood disorder from BD I or MDD, in terms of brain abnormalities, a number of previous mood episodes, seasonal aggravations, the circadian rhythm, depressive admixtures, and comorbidity [6,25-28]. The results of this study also show that BD II exhibits more discriminating characteristics than other major mood disorders. In particular, the superiority of the mood prediction algorithm based on the circadian rhythm suggests indirectly that BD II is likely to be affected by the disturbance of the circadian rhythm compared with mood episodes of other mood disorders. On the contrary, it is also possible that better predictions for BD II might be an artifact of the relatively well-balanced dataset for BP II and the fact that standard splits in random forests are not well-suited for the imbalanced classification. We will need further research for this possibility in the future.

Circadian disturbances have been reported in MDD including diurnal mood variation [29], core body temperature abnormality [30], changes in secretion of melatonin and cortisol [31], circadian rhythm alteration induced by antidepressants [32], and sleep-wake cycle disruption [31,33]. BD has even more robust results reported than MDD in relation to circadian rhythms, from an association with circadian gene variants [34], through sleep and circadian phenotypes [4,6,7], to the therapeutic approaches focusing on circadian rhythms and sleep [35]. Previous researchers have reported that seasonal variations in mood, behavior, and diurnal preference, and irregular bed-rise times, are closely related to BD, suggesting the importance of circadian rhythms in BD [36]. In particular, studies showed a closer association with seasonality in BD II than in BD I [37] and a greater chronotherapeutic effect in BD II [38].

Although the diagnosis of a mood disorder may be the same, clinical features vary from person to person. Automatically and passively recorded data from a diverse range of routine lives can directly or indirectly provide rich information reflecting each person’s psychiatric characteristics. Clinical symptoms are assessed primarily through interviews or psychiatric scales, which are dependent on reports from the patient or caregiver and are prone to recall errors and subjective bias. Continuously collected digital log data can provide a personalized upgrade to traditional clinical information. The mood prediction algorithm from this study provides a timely opportunity for the practical application of these data to treatment, especially for preventing acute mood episodes and managing daily mood states. The rapid development of information and communication technology (ICT) will present new therapeutic paradigm shifts for both clinicians and patients and help to fill the care gap in existing conventional treatments.

Mood follows a flow, so if patients record their own mood state every day, it can be a powerful predictor of future mood. However, recording daily mood requires ongoing attention and effort. As adherence is a key issue in mood disorder treatment, compliance would be improved if a patient’s condition could be managed and analyzed without any special effort or action. Using a smartphone and wearable device is a simple and convenient way to collect data to predict the mood state or pathological mood episodes.

Many studies have been conducted so far to predict mood or stress with data collection from smartphones such as the number of phone calls and text messages communicating with other people [16], entropy of subject’s location changes based on the global positioning system (GPS) [39], behavioral movement detection from accelerometer sensors [17,40], ambient light and noise sounds [17], and the paralinguistic feature of speech from smartphones [40]. LiKamWa et al used smartphone sensors to predict mood change in their study for 32 subjects and 2 months [16]. They analyzed the number and length of calls, short message service (SMS) text messages, and email communications; the usage number and pattern of apps; history of Web browser connections; and change of location information, reporting a prediction accuracy with 66%. After using the personalized prediction model, they could improve the accuracy up to 93%. Ma et al analyzed location information, user action and movement detection, ambient light and sounds, predicting the mood state with 50% accuracy in their study with 15 subjects and for 30 days [17]. One of the well-known projects for mood prediction study using smartphone-based sensors is the MONARCA project [41]. In the project, 12 actual BD were studied for 12 weeks. The mood prediction accuracy was achieved at 72% to 81% by using an accelerometer sensor and GPS-based location information. The prediction accuracy could be improved some more by including features of phone speech analysis. Gravenhorst et al also found from an extra study that higher use of social and entertainment apps was associated with lower stress and irritability [41]. Palmius et al could distinguish the mood depressive state from the nondepressive state with 85% accuracy by using features of GPS information including the entropy and circadian rhythm [42]. Carr et al studied if variability in phase and amplitude of the diurnal rhythm is related to variation of mood in bipolar and borderline personality disorder [43]. They investigated mood and diurnal variation for 4 days in 20 outpatients with BD, 14 with borderline personality disorder, and 20 healthy controls using a smartphone app, portable electrocardiogram, and actigraphy, reporting that for borderline personality disorder, there was a pattern of positive correlations between mood variability and variation in activity, sleep, and heart rate.

The previous studies were rather based on smartphone built-in sensors than on wearable devices. Smartphone sensors would be useful, but they have some limitations. First, smartphone is portable but not wearable. Therefore, even though smartphone is very easily applicable, it is not directly attached to the body so it is hard to collect data continuously without missing points over the timeline. Second, a privacy issue is serious. Many studies are depending on collecting data such as phone calls, SMS text messages, and GPS information, which are very sensitive and hard to be collected for a long time. However, a psychiatric study of mood prediction usually should need a long-time follow-up. The existing related studies have reported quite promising results in terms of mood prediction. However, they are mostly not analyzing actual BD patients but studying students or ordinary people without a mental disorder under the laboratory experiment setting or under an artificially instrumented environment. The number of analyzed people is limited to a small size and the length of study is not more than an annual period.

In contrast, our study was based on big data collection and analysis for about 2 years from 55 actual major mood disorder patients, which is reliably measured by a wearable device (a popular commercial product). To our knowledge, the proposed rhythm features (Multimedia Appendix 2) are unique and have never been tried before in model construction of the existing mood prediction studies. The accuracy of the proposed model performance is also reasonable compared with the existing related studies.

This study has several clinical strengths. First, we prospectively collected a vast amount of data for about 2 years from study subjects with mood disorders, accumulating 52,884 days of samples. Second, automatically passively recorded digital data from patients were collected using ICT, and the collected data were categorized, processed, and analyzed according to our hypotheses, regarding the influence of circadian rhythms, to obtain 130 daily features. As a result, from 260,390 to 6,874,920 total features (in the case of a complete dataset=2003 sample days×130 features and an incomplete dataset=52,884 sample days×30 features) were available for machine learning training and played an important role in improving predictability. Third, clinically significant prediction performance for a mood state or pathological mood episode was demonstrated using automatically recorded passive digital log data, in the absence of clinically derived mood information. Finally, the personalized prediction algorithm showed the potential to apply precision medicine principles to psychiatry.

Limitations

This study has some limitations. First, more intrinsic and preemptive genetic and biological assessments related to circadian rhythms were not included, as this would have required a more complex study design and analysis methods. We are planning to include these in future research. Second, the prediction capabilities were not high in some cases. However, the purpose of this study was not to diagnose mood disorders but to predict the mood state or episodes in patients with a diagnosed mood disorder, to improve the prognosis of patients with mood disorders by self-monitoring and self-care of moods and circadian rhythms in daily life through the mood prediction system developed in this study. Third, when analyzing the mood state, HAMS and LAMS were applied separately according to the absolute score of mood. This might not be an accurate reflection of the mood state, as we did not apply the high and low mood scores correctly. However, it was not easy to verify the accuracy of daily mood scores unlike the mood episodes, because the subjective mood score recorded by an individual may vary according to individual characteristics and tendencies. Finally, we reported the model performance results for individuals from our collected dataset only. Therefore, it is not sure how well the results will generalize to a new population. One way of addressing this properly would be by using cross-validation for different unseen patients. Some individuals would be selected for testing and others for training. The model would be trained using the training individuals and evaluated using the test individuals, and the procedure would be repeated for multiple splits as in the usual cross-validation. Thus, it is necessary for an additional future study to test and secure more external validity.

Conclusions

To our knowledge, this is the first study to develop a prediction system using only passive digital phenotypes from patients with mood disorders for a prolonged period of time. We have developed and verified mood state and pathological mood episode prediction algorithms using only automatically recorded passive data. On the basis of the results of this study, mood prediction algorithms can be applied therapeutically to improve clinical outcomes and the prognosis of patients with mood disorders. This study is just the first step toward future digital and precision medicine in the psychiatric field [44]. In future, a revolutionary change in psychiatric treatment will occur through the establishment of an integrated platform with genetic information and biological therapy.

Acknowledgments

This independent research study was supported by the Korea Health 21 R&D Project funded by the Ministry of Health & Welfare, Republic of Korea (HM14C2606 and HI14C3212), and the National Research Foundation of Korea (2016M3C7A1904345 and 2017M3A9F1031220). Most of all, we express our gratitude to the participants who took part in the study.

Authors' Contributions

CHC, TL, MGK, HPI, LK, and HJL designed and managed the study. CHC, TL, MGK, and HJL were responsible for data collection and trial management. CHC, TL, MGK, and HJL did the data analysis. CHC, TL, MGK, HPI, LK, and HJL provided expert advice on service, clinical, and patient-related matters. All authors contributed to writing and editing the paper. The first 2 authors contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic and clinical information of the subjects.

[DOCX File, 22KB - jmir_v21i4e11029_app1.docx]

Multimedia Appendix 2

The proposed basic features to check circadian rhythm from the automatically measured passive digital log data of patients with mood disorders.

[DOCX File, 17KB - jmir_v21i4e11029_app2.docx]

Multimedia Appendix 3

Different parameter setting affects model performance. The legend means the model test period q and the horizontal axis means the model training period p has 1 time of q days, 2 times of q days, 3 times of q days, and so on. The vertical axis means the area under the curve result after evaluating model performance with a parameter combination of p and q.

[DOCX File, 18KB - jmir_v21i4e11029_app3.docx]

Multimedia Appendix 4

The entire receiver operating characteristic curves of the main manuscript are presented here for clearer understanding of some unreported calibration in sensitivity and specificity.

[DOCX File, 194KB - jmir_v21i4e11029_app4.docx]

Multimedia Appendix 5

This is a Kernel Density Estimate plot describing variance of the model performance that was reported in the main manuscript (Mood state labeling with 50% cut-off) for the mood disorder group ALL. The horizontal axis presents the area under the curve (AUC) distribution and the vertical axis presents density of each AUC observation from multiple performance evaluation rounds. The total number of evaluation rounds is 668.

[DOCX File, 46KB - jmir_v21i4e11029_app5.docx]

Multimedia Appendix 6

The number of samples used in the mood state and mood episode prediction.

[DOCX File, 16KB - jmir_v21i4e11029_app6.docx]

Multimedia Appendix 7

Prediction performance comparison between personalized model versus general model.

[DOCX File, 14KB - jmir_v21i4e11029_app7.docx]

Multimedia Appendix 8

The area under the curve performance of each partially constructed model. The first column means a group of features that are missed in the model construction, and the second column is the number of features that are used in the model construction except the group of missing features. The rest of columns are the constructed model performance without using the missing features.

[DOCX File, 15KB - jmir_v21i4e11029_app8.docx]

References

1. World Health Organization: Geneva; 2008. The Global Burden of Disease 2004 Update URL: https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/ [accessed 2019-04-12] [WebCite Cache ID 77Zo9p2sj]
2. Simon GE. Social and economic burden of mood disorders. Biol Psychiatry 2003 Aug 1;54(3):208-215. [doi: 10.1016/S0006-3223(03)00420-7] [Medline: 12893097]

3. Malhi GS, Kuiper S. Chronobiology of mood disorders. Acta Psychiatr Scand Suppl 2013(444):2-15. [doi: 10.1111/acps.12173] [Medline: 23909692]
4. McClung CA. Circadian genes, rhythms and the biology of mood disorders. Pharmacol Ther 2007 May;114(2):222-232 [FREE Full text] [doi: 10.1016/j.pharmthera.2007.02.003] [Medline: 17395264]
5. Cho C, Lee H. Why do mania and suicide occur most often in the spring? Psychiatry Investig 2018 Mar;15(3):232-234 [FREE Full text] [doi: 10.30773/pi.2017.12.20] [Medline: 29566464]
6. Moon J, Cho C, Son GH, Geum D, Chung S, Kim H, et al. Advanced circadian phase in mania and delayed circadian phase in mixed mania and depression returned to normal after treatment of bipolar disorder. EBioMedicine 2016 Sep;11:285-295 [FREE Full text] [doi: 10.1016/j.ebiom.2016.08.019] [Medline: 27543154]
7. Cho C, Moon J, Yoon H, Kang S, Geum D, Son G, et al. Molecular circadian rhythm shift due to bright light exposure before bedtime is related to subthreshold bipolarity. Sci Rep 2016 Dec 22;6:31846 [FREE Full text] [doi: 10.1038/srep31846] [Medline: 27545669]
8. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May 12;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
9. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. J Am Med Assoc 2016 Dec 13;316(22):2353-2354. [doi: 10.1001/jama.2016.17438] [Medline: 27898975]
10. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. J Am Med Assoc 2016 Feb 9;315(6):551-552. [doi: 10.1001/jama.2015.18421] [Medline: 26864406]
11. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Washington, DC: American Psychiatric Pub; 2013.
12. Cho C, Ahn Y, Kim SJ, Ha TH, Jeon HJ, Cha B, et al. Design and methods of the Mood Disorder Cohort Research Consortium (MDCRC) study. Psychiatry Investig 2017 Jan;14(1):100-106 [FREE Full text] [doi: 10.4306/pi.2017.14.1.100] [Medline: 28096882]
13. Lee H, Lee H, Moon J, Lee T, Kim M, In H, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. Psychiatry Investig 2017 Mar;14(2):179-185 [FREE Full text] [doi: 10.4306/pi.2017.14.2.179] [Medline: 28326116]
14. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
15. Andrews S, Ellis DA, Shaw H, Piwek L. Beyond self-report: tools to compare estimated and real-world smartphone use. PLoS One 2015;10(10):e0139004 [FREE Full text] [doi: 10.1371/journal.pone.0139004] [Medline: 26509895]
16. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services. 2013 Presented at: MobiSys'13; Jun 25-28, 2013; Taipei, Taiwan p. 389-401. [doi: 10.1145/2462456.2464449]
17. Ma Y, Xu B, Bai Y, Sun G, Zhu R. Daily Mood Assessment Based on Mobile Phone Sensing. In: Proceedings of the 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: BSN'12; May 9-12, 2012; London, UK p. 142-147. [doi: 10.1109/BSN.2012.3]
18. Leo B. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
19. Ethem A. Introduction to Machine Learning. Cambridge, MA: The MIT Press; 2014.
20. 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 Feb 1;12:2825-2830 [FREE Full text]
21. LeGates TA, Fernandez DC, Hattar S. Light as a central modulator of circadian rhythms, sleep and affect. Nat Rev Neurosci 2014 Jul;15(7):443-454 [FREE Full text] [doi: 10.1038/nrn3743] [Medline: 24917305]
22. Grandin LD, Alloy LB, Abramson LY. The social zeitgeber theory, circadian rhythms, and mood disorders: review and evaluation. Clin Psychol Rev 2006 Oct;26(6):679-694. [doi: 10.1016/j.cpr.2006.07.001] [Medline: 16904251]
23. Boivin DB. Influence of sleep-wake and circadian rhythm disturbances in psychiatric disorders. J Psychiatry Neurosci 2000 Nov;25(5):446-458 [FREE Full text] [Medline: 11109296]
24. Greden JF. The burden of recurrent depression: causes, consequences, and future prospects. J Clin Psychiatry 2001;62(Suppl 22):5-9 [FREE Full text] [doi: 10.4236/jfrm.2014.31001] [Medline: 11599650]
25. Berk M, Dodd S. Bipolar II disorder: a review. Bipolar Disord 2005 Feb;7(1):11-21. [doi: 10.1111/j.1399-5618.2004.00152.x] [Medline: 15654928]
26. Heun R, Maier W. The distinction of bipolar II disorder from bipolar I and recurrent unipolar depression: results of a controlled family study. Acta Psychiatr Scand 1993 Apr;87(4):279-284. [Medline: 8488750]
27. Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, et al. A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. Arch Gen Psychiatry 2003 Mar;60(3):261-269. [doi: 10.1001/archpsyc.60.3.261] [Medline: 12622659]
28. Liu JX, Chen YS, Hsieh JC, Su TP, Yeh TC, Chen LF. Differences in white matter abnormalities between bipolar I and II disorders. J Affect Disord 2010 Dec;127(1-3):309-315. [doi: 10.1016/j.jad.2010.05.026] [Medline: 20598752]

29. Murray G. Diurnal mood variation in depression: a signal of disturbed circadian function? J Affect Disord 2007 Sep;102(1-3):47-53. [doi: 10.1016/j.jad.2006.12.001] [Medline: 17239958]
30. Goetze U, Tölle R. Circadian rhythm of free urinary cortisol, temperature and heart rate in endogenous depressives and under antidepressant therapy. Neuropsychobiology 1987;18(4):175-184. [doi: 10.1159/000118414] [Medline: 3454423]
31. Monteleone P, Maj M. The circadian basis of mood disorders: recent developments and treatment implications. Eur Neuropsychopharmacol 2008 Oct;18(10):701-711. [doi: 10.1016/j.euroneuro.2008.06.007] [Medline: 18662865]
32. Sprouse J, Braselton J, Reynolds L. Fluoxetine modulates the circadian biological clock via phase advances of suprachiasmatic nucleus neuronal firing. Biol Psychiatry 2006 Oct 15;60(8):896-899. [doi: 10.1016/j.biopsych.2006.03.003] [Medline: 16631132]
33. Casper RC, Redmond DE, Katz MM, Schaffer CB, Davis JM, Koslow SH. Somatic symptoms in primary affective disorder. Presence and relationship to the classification of depression. Arch Gen Psychiatry 1985 Nov;42(11):1098-1104. [doi: 10.1001/archpsyc.1985.01790340082012] [Medline: 3863548]
34. Gonzalez R. The relationship between bipolar disorder and biological rhythms. J Clin Psychiatry 2014 Apr;75(4):e323-e331. [doi: 10.4088/JCP.13r08507] [Medline: 24500332]
35. Frank E, Kupfer DJ, Thase ME, Mallinger AG, Swartz HA, Fagiolini AM, et al. Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. Arch Gen Psychiatry 2005 Sep;62(9):996-1004. [doi: 10.1001/archpsyc.62.9.996] [Medline: 16143731]
36. Geoffroy PA, Bellivier F, Scott J, Etain B. Seasonality and bipolar disorder: a systematic review, from admission rates to seasonality of symptoms. J Affect Disord 2014 Oct;168:210-223. [doi: 10.1016/j.jad.2014.07.002] [Medline: 25063960]
37. Friedman E, Gyulai L, Bhargava M, Landen M, Wisniewski S, Foris J, et al. Seasonal changes in clinical status in bipolar disorder: a prospective study in 1000 STEP-BD patients. Acta Psychiatr Scand 2006 Jun;113(6):510-517. [doi: 10.1111/j.1600-0447.2005.00701.x] [Medline: 16677228]
38. Krauss SS, Depue RA, Arbisi PA, Spoont M. Behavioral engagement level, variability, and diurnal rhythm as a function of bright light in bipolar II seasonal affective disorder: an exploratory study. Psychiatry Res 1992 Aug;43(2):147-160. [doi: 10.1016/0165-1781(92)90129-Q] [Medline: 1410070]
39. Madan A, Cebrian M, Lazer D, Pentland A. Social sensing for epidemiological behavior change. In: Proceedings of the 12th ACM international conference on Ubiquitous computing. 2010 Presented at: UbiComp'10; September 26-29, 2010; Copenhagen, Denmark p. 291-300. [doi: 10.1145/1864349.1864394]
40. Calvo RA, D'Mello S. Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans Affective Comput 2010 Jan;1(1):18-37. [doi: 10.1109/T-AFFC.2010.1]
41. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. Pers Ubiquit Comput 2014 Sep 21;19(2):335-353. [doi: 10.1007/s00779-014-0829-5]
42. Palmius N, Tsanas A, Saunders KE, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017 Dec;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
43. Carr O, Saunders KE, Tsanas A, Bilderbeck AC, Palmius N, Geddes JR, et al. Variability in phase and amplitude of diurnal rhythms is related to variation of mood in bipolar and borderline personality disorder. Sci Rep 2018 Dec 26;8(1):1649 [FREE Full text] [doi: 10.1038/s41598-018-19888-9] [Medline: 29374207]
44. Cho C, Lee H. Could digital therapeutics be a game changer in psychiatry? Psychiatry Investig 2019 Feb;16(2):97-98 [FREE Full text] [doi: 10.30773/pi.2019.01.20] [Medline: 30808114]

Abbreviations

AMS: absolute mood score
AUC: area under the curve
BD: bipolar disorder
BD I: bipolar disorder type 1
BD II: bipolar disorder type 2
DE: depressive episode
DSM: Diagnostic and Statistical Manual of Mental Disorders
GPS: global positioning system
HAMS: high absolute mood score
HME: hypomanic episode
ICT: information and communication technology
LAMS: low absolute mood score
MDCRC: Mood Disorder Cohort Research Consortium
MDD: major depressive disorder
ME: manic episode
NE: no episode

Edited by G Eysenbach; submitted 16.05.18; peer-reviewed by JH Baek, SA Strejilevich, E Moon, F Agakov; comments to author 07.01.19; revised version received 03.03.19; accepted 29.03.19; published 17.04.19.

Please cite as:
Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ
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;21(4):e11029
URL: http://www.jmir.org/2019/4/e11029/
doi:10.2196/11029
PMID:30994461

Original Paper

Validation of Two Automatic Blood Pressure Monitors With the Ability to Transfer Data via Bluetooth

Madeleine Wetterholm[1], MD; Stephanie Erika Bonn[1], PhD; Christina Alexandrou[1,2], MSc; Marie Löf[2], PhD; Ylva Trolle Lagerros[1,3], MD, PhD

[1]Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

[2]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[3]Obesity Center, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

Corresponding Author:
Madeleine Wetterholm, MD
Clinical Epidemiology Unit
Department of Medicine Solna
Karolinska Institutet
Eugeniahemmet T2
Stockholm, SE 171 76
Sweden
Phone: 46 851779183
Fax: 46 851779304
Email: madeleine.wetterholm@ki.se

Abstract

Background: Patients with chronic diseases are in need of regular health controls. Diabetes mellitus type 2 is currently the most prevalent chronic metabolic disease. A majority of diabetic patients have at least one comorbid chronic disease, where hypertension is the most common. The standard for blood pressure (BP) measurement is manual BP monitoring at health care clinics. Nevertheless, several advantages of self-measured BP have been documented. With BP data transfer from an automatic BP monitor via Bluetooth to software, for example, a smartphone app, home measurement could effectively be integrated into regular care.

Objective: The aim of this study was to validate two commercially available automatic BP monitors with the ability to transfer BP data via Bluetooth (Beurer BM 85 and Andersson Lifesense BDR 2.0), against manual BP monitoring in patients with type 2 diabetes.

Methods: A total of 181 participants with type 2 diabetes were recruited from 6 primary care centers in Stockholm, Sweden. BP was first measured using a manual BP monitor and then measured using the two automatic BP monitors. The mean differences between the automatic and manual measurements were calculated by subtracting the manual BP monitor measurement from the automatic monitor measurement. Validity of the two automatic BP monitors was further assessed using Spearman rank correlation coefficients and the Bland-Altman method.

Results: In total, 180 participants, 119 men and 61 women, were included. The mean age was 60.1 (SD 11.4) years and the mean body mass index was 30.4 (SD 5.4) kg/m^2. The mean difference between the Beurer BM 85 and the manual BP monitor was 11.1 (SD 11.2) mmHg for systolic blood pressure (SBP) and 8.0 (SD 8.1) mmHg for diastolic blood pressure (DBP). The mean difference between the Andersson Lifesense BDR 2.0 and the manual BP monitor was 3.2 (SD 10.8) mmHg for SBP and 4.2 (SD 7.2) mmHg for DBP. The automatic BP measurements were significantly correlated ($P<.001$) with the manual BP measurement values (Andersson Lifesense BDR 2.0: $r=0.78$ for SBP and $r=0.71$ for DBP; Beurer BM 85: $r=0.78$ for SBP and $r=0.69$ for DBP).

Conclusions: The two automatic BP monitors validated measure sufficiently accurate on a group level, with the Andersson Lifesense BDR 2.0 more often falling within the ranges for what is acceptable in clinical practice compared with the Beurer BM 85.

(J Med Internet Res 2019;21(4):e12772) doi:10.2196/12772

KEYWORDS

blood pressure monitors; diabetes mellitus, type 2; hypertension; methods; mHealth; self-care; self-management

Introduction

In recent years, there has been a rapid development in information and communication technology including the availability of digital devices for self-measurements and health promoting apps. Today, more and more patients are asking for digital solutions. Despite this, apps and devices for home measurement are seldom integrated into regular health care. Given the importance of self-care in the therapy of most chronic diseases, tools for self-measurement have the potential to play a greater role than they do today. Furthermore, the use of new technology such as mobile health, defined by the World Health Organization as medical or public health practice that is supported by mobile devices [1], could be a way to optimize care as a large number of patients can be reached at a lower cost.

Diabetes mellitus type 2 is currently the most prevalent chronic metabolic disease, and the number of cases is increasing worldwide. In 2015, the global prevalence of diabetes mellitus in the adult population (20-79 years) was 8.8%, and it is expected to rise to 10.4% by 2040 [2]. A majority of diabetic patients have at least one comorbid chronic disease of which hypertension, a powerful predictor of cardiovascular risk, is the most common [3,4].

Today, the standard method for diagnosis of hypertension as well as for blood pressure (BP) control in antihypertensive-treated patients is manual BP monitoring at health care clinics. Nevertheless, several advantages of self-measured BP have been documented [5]. For example, measurement at home can provide a more realistic appraisal of habitual BP than that can be obtained at a health care clinic, that is, eliminate the risk of the so-called *white coat hypertension* when the BP is high only in the clinical setting [6]. In addition, studies have also shown that the opposite, masked hypertension, that is, normal BP when measured at a clinic, but high BP when measured at home, is associated with an increased cardiovascular risk similar to the risk of patients with persistent hypertension [7,8]. As patients with diabetes have a high prevalence (47%) of masked hypertension [9], the need for validated automatic BP monitors in this specific population is of great importance.

Self-measurement at home could possibly improve patient adherence to both BP controls and treatment [10,11]. However, reporting of self-measured BP can be modified by the patients if the values for some reason do not seem suitable to them [12]. With automatic data transfer via Bluetooth to software, for example, a smartphone app, reporting bias as well as misreporting can be avoided. However, commercial automatic BP monitors are seldom validated, and to the best of our knowledge, no automatic BP monitor with data transfer via Bluetooth has been validated in patients with type 2 diabetes previously. Thus, in this study, we set out to validate two on the Swedish market commercially available automatic BP monitors (Beurer BM 85 Bluetooth and Andersson Lifesense BDR 2.0), with the ability to transfer data via Bluetooth, against manual BP monitoring in patients with type 2 diabetes.

Methods

Recruitment of Participants

This study was performed using BP data collected at baseline from all participants in the DiaCert-study, a randomized controlled trial of patients with type 2 diabetes. The study design has been described in detail previously [13]. A total of 181 participants were recruited from 6 primary care centers in Stockholm, Sweden. Inclusion criteria were as follows: being diagnosed with diabetes type 2, age above 18 years, being able to read and understand Swedish, being able to walk, and having access to and being able to use a smartphone. Overall, one participant did not have data on BP. Due to battery discharge or arm circumference larger than the recommended for the BP monitor cuffs, that is, more than 36 cm for Beurer BM 85 and more than 32 cm for Andersson Lifesense BDR 2.0, 11 participants did not have data from Beurer BM 85 and 25 participants did not have data from Andersson Lifesense BDR 2.0. In total, BP was measured using Beurer BM 85 in 169 participants and using Andersson Lifesense BDR 2.0 in 155 participants. All participants provided written consent before participating in the study. The study was approved by the Regional Ethical Review Board, Stockholm, Sweden (Dnr: 2016/2041-31/2; 2016/99-32; 2017/1406-32; 2018/286-32).

The Procedure

BP, weight, height, and waist circumference were measured by study personnel at the baseline meeting. This has been described in detail previously [13]. Smoking status (never, former, or current) was assessed through a questionnaire. The BP measurements were performed after at least 5 min of rest. Participants were seated with their legs uncrossed in a quiet room, and they were instructed to avoid talking during the procedure. The upper left arm of each participant was used for the BP measurement. BP was first measured once using the manual BP monitor and then measured once using both automatic BP monitors with no specific order.

The Automatic Monitors

The monitors Beurer BM 85 Bluetooth (Beurer GmbH. Ulm, Germany) and Andersson Lifesense BDR 2.0 (Guangdong Transtek Medical Electronics Co. Ltd. Zhongshan, China) are automatic devices for measuring BP at the upper arm. Both monitors can transfer data via Bluetooth to digital tools.

Beurer BM 85 has a pressure range of 0 to 300 mmHg and a memory capacity of 60 measurements for 2 users. It can calculate the average value of all saved measures as well as the average of morning and evening measurements during the last 7 days. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) as well as the heart rate are displayed on a liquid crystal digital (LCD) display. The monitor can identify an irregular heartbeat, which is then displayed with a symbol on the LCD screen. The included standard cuff for Beurer BM 85 is applicable to arm circumferences ranging from 22 to 36 cm. The dimensions of the device are 180×100×40 mm, and the weight of the monitor without the cuff is 317 grams. The Beurer BM 85 is equipped with a rechargeable lithium-ion battery (3.7 V/400 mAh) that has a battery life of approximately 50 measurements.

Andersson Lifesense BDR 2.0 has a pressure range of 0 to 300 mmHg and a memory capacity of 60 measurements for 2 users. It can calculate an average value of the last 3 measurements. SBP and DBP as well as the heart rate are displayed on an LCD display. The monitor can identify an irregular heartbeat, which is then displayed with a symbol on the LCD screen. The included standard cuff for Andersson Lifesense BDR 2.0 is applicable to arm circumferences ranging from 22 to 32 cm. The dimensions of the device are 180×99×40 mm, and the weight of the device without the cuff is 300 grams. For the Andersson Lifesense BDR 2.0, 4 AAA-size alkaline batteries are needed. The device has an approximate capacity of 300 measurements.

Statistical Analysis

We categorized the participants into low or high BP. Low BP was defined as SBP less than 140 mmHg and DBP less than 90 mmHg. High BP was defined as SBP 140 mmHg or above or DBP 90 mmHg or above. We used this classification as a BP of 140/90 mmHg is the diagnostic cut-off for the definition of hypertension in Europe [14]. Characteristics are presented as mean (SD) and n (%) for continuous and categorical variables, respectively. To assess whether there were any statistically significant differences between participants with low and high BP, Chi-square test was performed for categorical variables and 2-sided *t* tests were performed for continuous variables. Differences between the automatic monitors and manual BP measurements were calculated by subtracting the manual measurement from the automatic ones. Participants were divided into four categories classified by the differences according to whether they were within 5, 10, 15, or more than 15 mmHg. Separate variables were created for systolic and diastolic pressure. We conducted a sensitivity analysis to see if the result of Beurer BM 85 differed when including only the 155 participants in whom BP also was measured using Andersson Lifesense BDR 2.0. The Bland-Altman method was used to assess systematic differences in BP measurements between the manual and automatic monitors and as a graphical evaluation of the associations [15]. The difference in BP between the automatic monitor and the manual monitor was plotted on the y-axis and the mean of the two monitor measurements on the x-axis. The limits of agreement, equal to ±2SD of the mean difference, provide a measure of the variation. We assessed Spearman rank correlation coefficients between automatic and manual measurements to further examine the validity by measuring the degree of association. The significance level was set to .05. Analyses were performed using STATA 14 (Stata Corporation, College Station, TX, USA).

Results

This study included 180 participants (119 men and 61 women) with a mean age of 60.1 (SD 11.4) years. Characteristics of all participants and according to low (n=83) and high BP (n=97) are shown in Table 1. Participants with low BP and high BP did not differ significantly with respect to age, gender, or smoking status. However, there was a statistically significant difference in body mass index (BMI; *P*=.02) between the high and low BP groups with higher BMI in the high BP group. In addition, there was a statistically significant difference in waist circumference (*P*=.04) for men with greater waist circumference in the high BP group. The mean BP values for all participants with the manual monitor were 138 (SD 15.5) mmHg for SBP and 83 (SD 9.7) mmHg for DBP. The Beurer BM 85 and Andersson Lifesense BDR 2.0 mean BP values are shown in Table 1.

The mean difference between the Beurer BM 85 and the manual BP monitor was 11.1 (SD 11.2) for SBP and 8.0 (SD 8.1) for DBP. The mean difference between the Andersson Lifesense BDR 2.0 and the manual BP monitor was 3.2 (SD 10.8) for SBP and 4.2 (SD 7.2) for DBP. The number of measurements that differed from the manual measurements by 5, 10, 15 or less, and more than 15 mmHg are shown in Table 2. For Beurer BM 85, 49.1% (83/169) of all measurements differed by 10 mmHg or less in SBP and 30.8% (52/169) by 5 mmHg or less for DBP. For Andersson Lifesense BDR 2.0, 69.7% (108/155) of all measurements differed by 10 mmHg or less in SBP and 49.0% (76/155) by 5 mmHg or less for DBP. In sensitivity analysis, the results of Beurer BM 85 did not differ when including only the 155 participants in whom BP was measured using both automatic BP monitors (data not shown).

Table 1. Characteristics of study participants by categories of low and high blood pressure.

Variable	Total (N=180)	Low blood pressure; <140/<90 mmHg (n=83)	High blood pressure; ≥140/or ≥90 mmHg (n=97)	*P* value[a]
Age (years), mean (SD)	60.1 (11.4)	59.2 (11.2)	60.8 (11.5)	.36
Male, n (%)	119 (66.1)	56 (47.1)	63 (52.9)	.72
Body mass index (kg/m^2), mean (SD)	30.4 (5.4)	29.4 (5.1)	31.3 (5.4)	.02
Waist circumference (cm), mean (SD)				
All	107.5 (14.8)	106.3 (13.7)	108.5 (15.6)	.34
Male	110.1 (14.4)	107.3 (12.7)	112.7 (15.4)	.04
Female	102.3 (14.2)	104.3 (15.8)	100.7 (12.8)	.33
Smoking status, n (%) .54				
Never	72 (40.0)	35 (42.2)	37 (38.1)	
Former	71 (39.4)	30 (36.1)	41 (42.3)	
Current	20 (11.1)	11 (13.3)	9 (9.3)	
Missing	17 (9.4)	7 (8.4)	10 (10.3)	
Systolic blood pressure (mmHg), mean (SD)				
Manual	138 (15.5)	126 (8.7)	148 (13.0)	<.001
Beurer BM 85[b]	149 (18.2)	137 (12.2)	159 (15.8)	<.001
Andersson Lifesense BDR 2.0[c]	140 (18.5)	129 (11.0)	151 (17.5)	<.001
Diastolic blood pressure (mmHg), mean (SD)				
Manual	83 (9.7)	78 (6.3)	88 (9.6)	<.001
Beurer BM 85[b]	91 (10.1)	87 (7.8)	94 (10.6)	<.001
Andersson Lifesense BDR 2.0[c]	87 (10.2)	82 (7.1)	91 (10.4)	<.001

[a] *P* value from Chi-square test or *t* test between groups of low blood pressure and high blood pressure.

[b] n=169.

[c] n=155.

Table 2. Validation results of the Beurer BM 85 and Andersson Lifesense BDR 2.0 (the number of measurements that differed from the manual blood pressure measurement by 5, 10, 15 or less, and more than 15 mmHg).

Variable	≤5 mmHg	≤10 mmHg	≤15 mmHg	>15 mmHg
Beurer BM 85				
SBP[a], n (%)	47 (27.8)	83 (49.1)	109 (64.5)	60 (35.5)
DBP[b], n (%)	52 (30.8)	105 (62.1)	146 (86.4)	23 (13.6)
Andersson Lifesense BDR 2.0				
SBP, n (%)	68 (43.9)	108 (69.7)	129 (83.2)	26 (16.8)
DBP, n (%)	76 (49.0)	123 (79.4)	147 (94.8)	8 (5.2)

[a] SBP: systolic blood pressure.

[b] DBP: diastolic blood pressure.

The differences in SBP and DBP in relation to the mean between the automatic and the manual monitors are shown in Bland-Altman plots in Figures 1 and 2. Most data points fall within the limits of agreement (±2SD), although it should be noted that the limits of agreement were wide and individual differences are shown. The data points in all Bland-Altman plots show a consistent horizontal pattern around the mean of the y-axis, that is, no trend was identified, that is, the accuracy did not seem to be impacted by the level of SBP or DBP. All of the automatic BP measurements were significantly correlated with the manual measurements. The Spearman correlation coefficient was *r*=0.78 for SBP and *r*=0.71 for DBP for Andersson Lifesense BDR 2.0, and *r*=0.78 for SBP and *r*=0.69 for DBP for Beurer BM 85 (*P*<.001 for all).

Figure 1. Bland-Altman plots of the differences between the Beurer BM 85 measurements and the manual measurements for systolic blood pressure (a) and diastolic blood pressure (b). The difference in blood pressure between Beurer BM 85 and the manual monitor is plotted on the y-axis and the mean of the two monitor measurements on the x-axis. Each data point represents one participant (n=169).

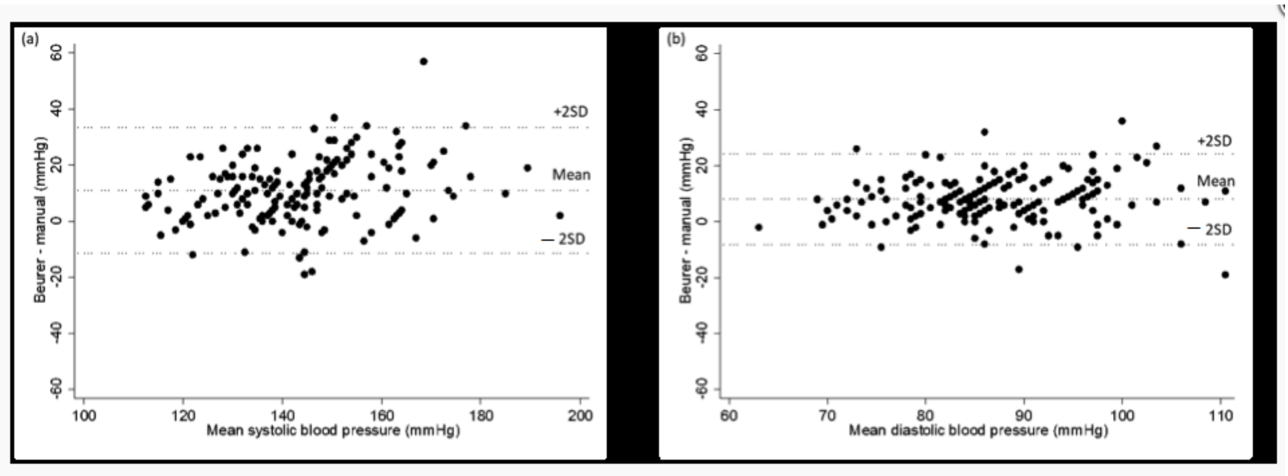
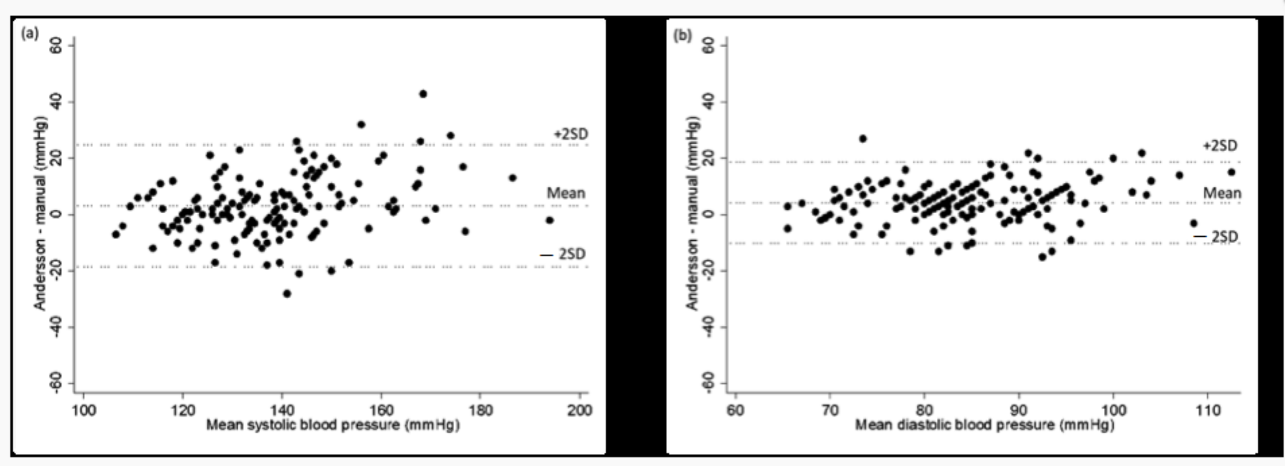


Figure 2. Bland-Altman plots of the differences between the Andersson Lifesense BDR 2.0 measurements and the manual measurements for systolic blood pressure (a) and diastolic blood pressure (b). The difference in blood pressure between the Andersson Lifesense BDR 2.0 and the manual monitor is plotted on the y-axis and the mean of the two monitor measurements on the x-axis. Each data point represents one participant (n=155).



Discussion

Principal Findings

The results of our study demonstrate that the mean difference between the manual and the automatic BP monitors Andersson Lifesense BDR 2.0 and Beurer BM 85 was small on a group level. The differences between the manual and automatic measurements were larger on an individual level.

There is a lack of studies validating automatic BP monitors in patients with type 2 diabetes. However, Masding et al [16] compared automatic home BP measurement and manual BP measurement with a previously validated 24-hour ambulatory BP monitor, which measured BP every 30 min during the day and every 60 min during the night, in 55 patients with type 2 diabetes. They found automatic home-measured BP superior to clinically measured BP. The mean difference between the automatic BP and the 24-hour ambulatory BP was 8.2 and 3.7 mmHg for SBP and DBP, respectively. The manual BP monitor compared with the 24-hour ambulatory BP monitor showed a mean difference of 10.9 and 3.8 mmHg for SBP and DBP, respectively. In our study, Andersson Lifesense BDR 2.0 showed a lower mean difference in SBP (3.2 mmHg).

Masding et al [16] have also predefined ranges for differences in BP that would be acceptable in clinical practice to be 10 mmHg for SBP and 5 mmHg for DBP. Comparing the results from our study, the Andersson Lifesense BDR 2.0 falls within these ranges more often than the Beurer BM 85. Although 69.7% (108/155) of automatic measurements performed using the Anderson Lifesense BDR 2.0 was within 10 mmHg from the manual measurement for SBP, only 49.1% (83/169) of automatic measurements with Beurer BM 85 were within this limit. For DBP, 49.0% (76/155) and 30.8% (52/169) were within 5 mmHg of difference for Andersson Lifesense BDR 2.0 and Beurer BM 85, respectively. On a group level, Andersson Lifesense BDR 2.0 meets the clinical ranges with a mean difference of 3.2 mmHg for SBP and 4.2 mmHg for DBP. Although their patient group was similar to ours, it should be noted that the studies differ in a number of aspects. First, in their study, BP was measured with the manual BP monitor at 3 different visits to the health clinic. Thereafter, the patients were instructed to use the automatic BP monitor at home at 3 specified times on 4 consecutive days, comparing it with the 24-hour ambulatory

BP monitor [16]. In our study, BP was measured manually and by automatic monitors on 1 occasion only.

Several automatic BP monitors without a Bluetooth connection have been validated in the general population [17-21]. Takahashi et al [17] validated three different automatic BP monitors against manual BP monitoring. All three of them showed better accuracy than the two monitors we validated in this study. In our study, the percentage of measurements that differed from the control measurements by 5, 10, and 15 mmHg or less for the monitor with the best accuracy, Andersson Lifesense BDR 2.0, were 43.9% (68/155), 69.7% (108/155), and 83.2% (129/155) for SBP and 49.0% (76/155), 79.4% (123/155), and 94.8% (147/155) for DBP. For one of the monitors validated by Takahashi et al, the corresponding numbers were 72%, 92%, and 98% for SBP and 82%, 98%, 100% for DBP. However, they only included 33 participants from the general population and BP was measured with each of the monitors at 3 times on each participant, giving a total of 99 measurements [17]. These differences make it difficult to compare results across studies. Furthermore, in our study, the automatic BP measurements were significantly, with correlation coefficients of .69 or higher, correlated with the manual BP measurements. However, to the best of our knowledge, no correlation analysis has been conducted in any validation study of automatic BP monitors compared with manual BP measurement.

The large sample size is a notable strength of our study. Furthermore, the participants were recruited from 6 primary care centers located in different areas with diverse populations and levels of socioeconomic status. With a mean age of 60 years, the participants are younger than the general patients with type 2 diabetes in Sweden (68 years old) [22], which may be due to the inclusion criteria of having a smartphone. However, 8 of 10 Swedes have a smartphone [23]. Our study also includes a larger number of men compared with women. This may primarily be a reflection of the higher prevalence of diabetes type 2 among men compared with women in Sweden [24,25]. Nevertheless, this study includes more women than any previous studies validating automatic BP monitors [16,17].

There is a lack of automatic BP monitors validated in the large population of patients with type 2 diabetes who are likely to use automatic BP monitors in practice. Not only are patients with diabetes often familiar with digital devices that help them with better management, for example, devices for self-monitoring of blood glucose, they are also in need of controls for comorbidities, such as hypertension, on a regular basis. Today, the routine practice for screening of hypertension consists of multiple visits to the health care clinic for repeated BP measurement. Including home measurements in the decision making would not only be more cost-efficient but also provide more reliable measurements with white coat hypertension and masked hypertension in mind. The fact that BP measurements in our study were performed in a clinical setting and not in a home setting, and only once using each monitor, may be a limitation of our study. Though, as the aim was to validate the automatic monitors against the standard method for BP measurement, the manual and the automatic measurements were performed on the same occasion at a health care clinic.

Digital solutions could meet the needs and expectations of patients by making health care more accessible. Validated devices also ensure the quality of the care. However, if the BP monitors’ ability to transfer BP data via Bluetooth is to be used in future health care, it is important to emphasize that transfer of data to, for example, an electronic health record has to be compliant with security regulations and without the risk of privacy invasion of stored health data on the cloud. In addition, offering tech support, for example, an in-app chat-support, for various issues such as data transfer and uploading data could potentially increase consumer confidence in digital self-measurement devices. Gamification techniques such as visual feedback messages or a platform allowing communication could have potential to further increase the motivation of the patients to engage in their own health care [26]. For future patients, the BP data together with other variables could possibly allow for personalized lifestyle recommendations.

Conclusions

In conclusion, no previous studies have validated automatic BP monitors with a Bluetooth connection in patients with type 2 diabetes. Our study shows that although the difference between the manual and the automatic BP monitors was greater on an individual level, the monitors are sufficiently accurate on a group level. Moreover, the Andersson Lifesense BDR 2.0 is more often falling within the BP ranges for what is acceptable in clinical practice compared with the Beurer BM 85.

Acknowledgments

The authors would like to acknowledge the personnel at participating primary care centers for the help in recruitment of study participants and, in particular, Klara Wiklander, Julia Ragnö, Kristin Hjörleifsdottir Steiner, and Monica Flodin. This study was supported by Stockholm County Council and Karolinska Institutet (4D) as well as through grants from the Swedish Research Council for Health, Working life, and Welfare (Dnr: 2016–00985) and funding from the strategic research area in care sciences (SFO-V) to SEB. The funders had no role in the design, analysis, or writing of this study.

Conflicts of Interest

None declared.

References

1. World Health Organization (WHO). World Health Organization. 2011. mHealth: New horizons for health through mobile technologies: second global survey on eHealth URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2019-02-25] [WebCite Cache ID 76Rke6A9v]
2. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
3. Pantalone KM, Hobbs TM, Wells BJ, Kong SX, Kattan MW, Bouchard J, et al. Clinical characteristics, complications, comorbidities and treatment patterns among patients with type 2 diabetes mellitus in a large integrated health system. BMJ Open Diabetes Res Care 2015;3(1):e000093 [FREE Full text] [doi: 10.1136/bmjdrc-2015-000093] [Medline: 26217493]
4. Hermans MP, Dath N. Prevalence and co-prevalence of comorbidities in Belgian patients with type 2 diabetes mellitus: a transversal, descriptive study. Acta Clin Belg 2018 Feb;73(1):68-74. [doi: 10.1080/17843286.2017.1348710] [Medline: 28679322]
5. Parati G, Stergiou GS, Asmar R, Bilo G, de Leeuw P, Imai Y, ESH Working Group on Blood Pressure Monitoring. European Society of Hypertension guidelines for blood pressure monitoring at home: a summary report of the Second International Consensus Conference on Home Blood Pressure Monitoring. J Hypertens 2008 Aug;26(8):1505-1526. [doi: 10.1097/HJH.0b013e328308da66] [Medline: 18622223]
6. Pickering TG, James GD, Boddie C, Harshfield GA, Blank S, Laragh JH. How common is white coat hypertension? J Am Med Assoc 1988 Jan 08;259(2):225-228. [Medline: 3336140]
7. Sega R, Trocino G, Lanzarotti A, Carugo S, Cesana G, Schiavina R, et al. Alterations of cardiac structure in patients with isolated office, ambulatory, or home hypertension: data from the general population (Pressione Arteriose Monitorate E Loro Associazioni [PAMELA] Study). Circulation 2001 Sep 18;104(12):1385-1392. [Medline: 11560854]
8. Fagard RH, Cornelissen VA. Incidence of cardiovascular events in white-coat, masked and sustained hypertension versus true normotension: a meta-analysis. J Hypertens 2007 Nov;25(11):2193-2198. [doi: 10.1097/HJH.0b013e3282ef6185] [Medline: 17921809]
9. Eguchi K, Ishikawa J, Hoshide S, Pickering TG, Shimada K, Kario K. Masked hypertension in diabetes mellitus: a potential risk. J Clin Hypertens (Greenwich) 2007 Aug;9(8):601-607 [FREE Full text] [Medline: 17673881]
10. Edmonds D, Foerster E, Groth H, Greminger P, Siegenthaler W, Vetter W. Does self-measurement of blood pressure improve patient compliance in hypertension? J Hypertens Suppl 1985 Apr;3(1):S31-S34. [Medline: 3870467]
11. Ogedegbe G, Schoenthaler A. A systematic review of the effects of home blood pressure monitoring on medication adherence. J Clin Hypertens (Greenwich) 2006 Mar;8(3):174-180 [FREE Full text] [Medline: 16522994]
12. Mengden T, Hernandez Medina RM, Beltran B, Alvarez E, Kraft K, Vetter H. Reliability of reporting self-measured blood pressure values by hypertensive patients. Am J Hypertens 1998 Dec;11(12):1413-1417. [Medline: 9880121]
13. Bonn SE, Alexandrou C, Hjörleifsdottir Steiner K, Wiklander K, Östenson CG, Löf M, et al. App-technology to increase physical activity among patients with diabetes type 2-the DiaCert-study, a randomized controlled trial. BMC Public Health 2018 Dec 10;18(1):119 [FREE Full text] [doi: 10.1186/s12889-018-5026-4] [Medline: 29316905]
14. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Böhm M, Task Force for the Management of Arterial Hypertension of the European Society of Hypertensionthe European Society of Cardiology. 2013 ESH/ESC practice guidelines for the management of arterial hypertension. Blood Press 2014 Feb;23(1):3-16. [doi: 10.3109/08037051.2014.868629] [Medline: 24359485]
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
16. Masding MG, Jones JR, Bartley E, Sandeman DD. Assessment of blood pressure in patients with Type 2 diabetes: comparison between home blood pressure monitoring, clinic blood pressure measurement and 24-h ambulatory blood pressure monitoring. Diabet Med 2001 Jun;18(6):431-437. [Medline: 11472460]
17. Takahashi H, Yoshika M, Yokoi T. Validation of three automatic devices for the self-measurement of blood pressure according to the European Society of Hypertension International Protocol revision 2010: the Omron HEM-7130, HEM-7320F, and HEM-7500F. Blood Press Monit 2015 Apr;20(2):92-97. [doi: 10.1097/MBP.0000000000000096] [Medline: 25462531]
18. Takahashi H, Yoshika M, Yokoi T. Validation of two automatic devices: Omron HEM-7252G-HP and Omron HEM-7251G for self-measurement of blood pressure according to the European Society of Hypertension International Protocol revision 2010. Blood Press Monit 2015 Oct;20(5):286-290. [doi: 10.1097/MBP.0000000000000127] [Medline: 25932887]
19. Guo YF, Zhang XQ, Fu JL, Zhang X, Shan HZ, Shi N, et al. Validation of the G.LAB MD2680 digital automatic blood pressure monitor according to international protocols. Blood Press Monit 2017 Aug;22(4):221-225. [doi: 10.1097/MBP.0000000000000254] [Medline: 28288006]
20. Chen W, Zeng ZL, Bing S, Li LY, Wang R, Wan Y. Validation of the Grandway MD2301 digital automatic blood pressure monitor according to the European Society of Hypertension International Protocol. Blood Press Monit 2016 Aug;21(4):259-261. [doi: 10.1097/MBP.0000000000000191] [Medline: 27093703]
21. Topouchian J, Agnoletti D, Blacher J, Youssef A, Ibanez I, Khabouth J, et al. Validation of four automatic devices for self-measurement of blood pressure according to the international protocol of the European Society of Hypertension. Vasc Health Risk Manag 2011;7:709-717 [FREE Full text] [doi: 10.2147/VHRM.S27193] [Medline: 22174581]

22. Gudbjörnsdottir S, Svensson A, Eliasson B, Eeg OK, Linder E, Samuelsson P, et al. [National Diabetes Register]. [Annual Report] URL: https://www.ndr.nu/pdfs/Arsrapport_NDR_2017.pdf [accessed 2018-09-26] [WebCite Cache ID 72ifG8iJW]
23. Davidsson P, Thoresson A. Svenskarna och internet–Undersökning om svenskarnas internetvanor [Swedes and internet]. 2017. Internetstiftelsen i Sverige [The Swedes and the Internet] URL: https://www.iis.se/docs/Svenskarna_och_internet_2017.pdf [accessed 2018-08-31] [WebCite Cache ID 725K1Eja6] URL: https://2017.svenskarnaochinternet.se/the-swedes-and-the-internet-2017-summary/ [accessed 2018-08-31] [WebCite Cache ID 725KUaqpe]
24. Wändell PE, Gåfvels C. Patients with type 2 diabetes aged 35-64 years at four primary health care centres in Stockholm County, Sweden. Prevalence and complications in relation to gender and socio-economic status. Diabetes Res Clin Pract 2004 Mar;63(3):195-203. [Medline: 14757291]
25. Carlsson AC, Wändell PE, Hedlund E, Walldius G, Nordqvist T, Jungner I, et al. Country of birth-specific and gender differences in prevalence of diabetes in Sweden. Diabetes Res Clin Pract 2013 Jun;100(3):404-408. [doi: 10.1016/j.diabres.2013.03.014] [Medline: 23618551]
26. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Dec;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

Abbreviations

BMI: body mass index
BP: blood pressure
DBP: diastolic blood pressure
LCD: liquid crystal digital
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 15.11.18; peer-reviewed by K Anderson, M Lozano-Lozano, B Chaudhry; comments to author 16.01.19; revised version received 29.01.19; accepted 30.01.19; published 17.04.19.

Please cite as:
Wetterholm M, Bonn SE, Alexandrou C, Löf M, Trolle Lagerros Y
Validation of Two Automatic Blood Pressure Monitors With the Ability to Transfer Data via Bluetooth
J Med Internet Res 2019;21(4):e12772
URL: https://www.jmir.org/2019/4/e12772/
doi:10.2196/12772
PMID:30994459

Original Paper

To Each Stress Its Own Screen: A Cross-Sectional Survey of the Patterns of Stress and Various Screen Uses in Relation to Self-Admitted Screen Addiction

Najmeh Khalili-Mahani[1,2], PhD, MEng, MA; Anna Smyrnova[3], BSc; Lisa Kakinami[3], PhD

[1]PERFORM Centre, Design and Computation Arts, Concordia University, Montreal, QC, Canada
[2]McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[3]PERFORM Centre, Department of Mathematics and Statistics, Concordia University, Montreal, QC, Canada

Corresponding Author:
Najmeh Khalili-Mahani, PhD, MEng, MA
PERFORM Centre
Design and Computation Arts
Concordia University
7200 Rue Sherbrooke O
Montreal, QC, H4B1R6
Canada
Phone: 1 5148482424 ext 5370
Email: najmeh.khalili-mahani@concordia.ca

Abstract

Background: The relationship between stress and screen addiction is often studied by exploring a single aspect of screen-related behavior in terms of maladaptive dependency or the risks associated with the content. Generally, little attention is given to the pattern of using different screens for different types of stressors, and variations arising from the subjective perception of stress and screen addiction are often neglected. Given that both addiction and stress are complex and multidimensional factors, we performed a multivariate analysis of the link between individual’s subjective perceptions of screen addiction, various types of stress, and the pattern of screen usage.

Objective: Using the media-repertoires framework to study usage patterns, we explored (1) the relation between subjective and quantitative assessments of stress and screen addiction; and (2) differences in stress types in relation to subjective screen addiction and different types of needs for screens. We hypothesized that interindividual heterogeneity in screen-related behavior would reflect coping differences in dealing with different stressors.

Methods: A multifactorial Web-based survey was conducted to gather data about screen-related behaviors (such as screen time, internet addiction, and salience of different types of screens and related activities), and different sources of stress (emotional states, perceptual risks, health problems, and general life domain satisfaction). We performed group comparisons based on whether participants reported themselves as addicted to internet and games (A1) or not (A0), and whether they had experienced a major life stress (S1) or not (S0).

Results: Complete responses were obtained in 459 out of 654 survey responders, with the majority in the S1A0 (44.6%, 205/459) group, followed by S0A0 (25.9%, 119/459), S1A1 (19.8%, 91/459), and S0A1 (9.5%, 44/459). The S1A1 group was significantly different from S0A0 in all types of stress, internet overuse, and screen time (*P*<.001). Groups did not differ in rating screens important for short message service (SMS) or mail, searching information, shopping, and following the news, but a greater majority of A1 depended on screens for entertainment (χ^2_3=20.5; *P*<.001), gaming (χ^2_3=35.6; *P*<.001), and social networking (χ^2_3=26.5; *P*<.001). Those who depended on screens for entertainment and social networking had up to 19% more emotional stress and up to 14% more perceptual stress. In contrast, those who relied on screens for work and professional networking had up to 10% higher levels of life satisfaction. Regression models including age, gender, and 4 stress types explained less than 30% of variation in internet use and less than 24% of the likelihood of being screen addicted.

Conclusions: We showed a robust but heterogeneous link between screen dependency and emotional and perceptual stressors that shift the pattern of screen usage toward entertainment and social networking. Our findings underline the potential of using ludic and interactive apps for intervention against stress.

(J Med Internet Res 2019;21(4):e11485) doi:10.2196/11485

KEYWORDS

psychological stress; addictive behavior; coping behavior; social network; gaming, internet; communications; telemedicine; eHealth

Introduction

Background

In *Understanding Media: The Extension of Man* [1] , Marshal McLuhan talked about the relation between media and stress by drawing attention to the work of the father of stress biology, Hans Selye, and the author of *The Stress of Life* [2] . Hans Selye had shown that the biological response to psychological threats was identical to the one caused by illness or predatory threat [3]. He called this generalized physiological response stress. A long history of research has ensued to show that although stress is an important adaptive response, chronic exposure to stress would cause various health problems [4]. Referring to the psychobiological evidence, McLuhan proposed that by the virtue of its speed in affective and cognitive stimulation, the electronic media could generate pervasive and cascade neurophysiological interactions that are similar to *stress* (as defined by Selye) would impact the brain's information processing capacity, leading to pleasure, fear, action, and memory. Today, with the ubiquity of mobile, computationally *smart*, globally connected and socially networked media, McLuhan's concern about the relation between media and stress adaptation becomes more relevant. Indeed, the relation between screens and stress is extensively researched. Researchers either investigate extremely problematic cases, in which screen addiction becomes a chronic stressor due to disruption of social, mental, and physical health, or investigate the role of screens in coping with chronic stress. Given the ubiquity of various forms of information and communication technologies (ICTs), and their high penetration in the industrial societies, many health researchers and industrial developers are exploring ways to innovate media-based interventions that can mitigate stress [5-9]. However, for the digital health applications (also known as telemedicine or eHealth) to be effective, they must first disambiguate and address the growing concerns about potential risks of screen addiction [10-12].

Screens and Coping With Psychological Stress

Numerous studies over decades have emphasized the impact of perception, appraisal, and internal and external factors that define an individual's risk of developing emotional or health problems due to stress. As screens are communication tools, they can alter the perceptual context and the cognitive and emotional appraisal processes through their messaging. To cope with stress successfully, individuals must dynamically change their perceptual and appraisal landscapes in order to activate cognitive and behavioral adaptations needed to control their stress [13-16]. Stress adaptation is a contextual process and individuals cope with stress through a mix of avoidance-approach [17] or problem- and emotion-focused coping strategies [18], based on their history of exposure and with different resilience factors [19]. In their most popular application, screens are either entertainment devices (that can facilitate avoidance-coping by providing distraction from the source of stress, and relaxation, through endless possibilities for playing games or watching videos), or electronic information centers (that can facilitate information-based and problem-focused approach-coping). The informatics and hypertextual nature of modern screens allows one to personalize them to their coping style according to their cultural or psychological needs. As new screens are also interactive, they not only provide communication and community but also facilitate anonymity and protection from the risks and anxieties of actual encounters. Mobile and internet-connected screens bridge temporal and spatial distances and extend possibilities of seeking support from social and professional networks in addressing specific or general problems. Portable and connected screens (laptops, tablets, and mobile phones) provide a wide range of search tools, production and computation software, and entertainment and playful options that extend an individual's sense of control not only over space and time but also across emotional and cognitive domains. As such, screens can also help with problem-focused coping to enhance confidence, control, or outcome prediction, for example, by extending one's functional resources through instantaneous and almost ubiquitous educational, information, communication, and computational resources.

It has been shown that following a major life event, about 57% of a general population (in the Netherlands, with 94% internet penetration) would use the internet for coping, specifically by playing games for mental disengagement and searching information [20]. Communication via blogging has been shown to reduce stress by increasing possibilities for emotion-focused and problem-focused coping through social support [21]. Internet-supported educational or health care interventions are viable for treating stressful physical and mental health conditions [22-31]. For older adults, who face a number of stressors including the loss of cognitive and executive agility, reduced mobility, and diminished social interaction, the opportunity to play video games [32-34] or to engage via online social networks [35,36] has shown positive cognitive and emotional benefits. A systemic review of over 5400 studies of mental health apps on *smartphones* (ie, mobile phones with augmented processing units, with a touch-screen, able to connect to the internet and equipped with accessories such as cameras, voice recorders, etc) suggests promising potential for this mode of intervention in depression and anxiety disorders [37]. There is even experimental evidence to show that being connected to social media can mitigate the physiological response to a psychosocially stressful condition [38] or that adding social media interventions may increase the therapeutic efficacy of pharmacological interventions in treatment of depression [39].

Stress-Related Risk of Screen Addiction

If screens can help an individual cope with stress, then it is also plausible that chronic stress would increase the risk of developing neurobiologically consequential screen addiction

[40]. The earliest clinical studies of screen addiction go back to the television era [41,42], followed by computer and video games [43], the internet [44-46], and more recently, mobile phone [47]. The target for the majority of these studies is young children or adolescents, or individuals, who suffer a quantifiable disruption in normal life domains (eg, health, finance, family, social relations, and work) as a result of compulsive usage of one technology. These studies underline the correlation between screen usage and stress-related psychopathology [48,49], or the negative health impacts of addiction to television [41,50], computer games [51,52], the internet [53,54], and social media [55-57]. Significant associations between problematic screen use and stressors such as familial instability [58-60] and parental styles [61,62], socioeconomic status and work load [63], have been reported. In a 1-year cohort study of more than 4160 young adults, moderate to excessive computer usage was associated with sleep disturbance in both men and women—whereas greater email/chat usage was correlated with greater risks of mental health problems in women, it was associated with lower perceived stress in men [64]. However, similar studies in the older and nonclinical population are still rare.

The Research Question

One of the current shortcomings in our knowledge of the relation between screens and stress is that the clinical classifications of screen addiction generally draw on 6 quantifying factors used to diagnose drug dependency: salience, tolerance, withdrawal, interpersonal conflicts, mood alterations, and relapse. However, it has been shown that a general clinical criterion of internet or gaming addiction ignores significant heterogeneity in the accessibility and the content of the medium to which one becomes addicted [65]. Variations in gender and age in terms of vulnerability to stress-related screen usage and self-evaluation of addiction are also important considerations [66,67].

A similar limitation exists in quantifying stress. There are numerous psychometric scales that estimate the risk of being stressed by considering combinations of the emotional and autonomic experience of distress, for example, perceived threats and anxiety [68,69], or life satisfaction [70], and perceptions of self-efficacy and control [71]. Although these questionnaires have common components that underline the stress psychobiology, they do not account for many individual or societal factors that influence the subjective stressfulness of a situation and modulate the functional reserves that are available to the individual for coping with daily stress. Although the clinical questionnaires are designed to be sensitive enough to diagnose the *problematic* or at-risk cases, they may not be sensitive to detecting subtle interindividual heterogeneities that explain variations in general daily screen usage for dealing with normal stressors of life. In the same vein, although there are strong objective markers to link addiction to neurobiology, the less explored individual and socio-relational components may better explain the likelihood of developing stress-related addictions to both drug and certain behaviors—eating, gambling, compulsive internet use, etc [72]. In fact, some argue against the pathological conceptualization of addiction as a purely biological phenomenon and emphasize the primacy of the individual's choice in seeking pleasure through repetition of a behavior [73].

The aim of this study is to explore the question of stress and screen addiction in a multi-factorial mixed-method fashion that allows us to examine the complexity of stress-related screen dependency.

Research Approach

In studying the behavioral and contextual differences in usage of communication technologies, media scholars suggest a repertoire-oriented framework that emphasizes the interrelation between different available technologies and underline the importance of characterizing the individual's choice in the amount of use of different media or content [74]. Existing studies of screen addiction narrowly focus on extreme abnormalities by comparing stratified demographics in relation with specific addictions (eg, gaming, gambling, social networking, and compulsive internet use) and specific clinical manifestations (eg, violence, attention deficit, depression, and anxiety disorders). The repertoires-oriented framework acknowledges the user's choice between different technologies and in the context of our research asks to what extent would the explanatory factors (in our case, stress or addiction) influence the *pattern* s of different screen uses? This pattern approach is particularly useful in studying the heterogeneity of screen usage arising from subjective versus objective assessments of stress or addiction. It has been shown that the *objective* quantification of stress (be it in terms of socioeconomic, psychometric, or other ratings) does not necessarily correspond to subjective perception of stress [75-79]. Similarly, the majority of definitions of addiction converge on the following elements: hedonic experience following engagement in the behavior, preoccupation with the behavior, loss of control, and suffering negative consequences as a result of losing oneself in the behavior [80]. Engaging in excessive computer use (for research, work, communication, playing, or relaxing) is not necessarily perceived as an addiction to those who engage in the activity. Are there common emotional, perceptual, health, and life domain stressors that distinguish those who consider themselves screen addicted? Do self-described screen addicts have higher scores of internet addiction and screen time? Do they differ from nonaddicts in evaluation of the importance of, access to, and dependence on different screen activities? And finally, are there subtypes of stress that would explain the self-rated screen addiction or the dependence on a given application of screens?

In this study, we have taken a repertoire-oriented approach [74] to explore the relation between stress and patterns of screen usage based on the individual's subjective assessment of stress and self-rated degree of screen addiction. We hypothesized that individuals who consider themselves screen addicted have higher stress levels than the nonaddicted and that there is a correlation between different types of stress and different types of screen usage to suggest individualized approaches for coping with stress via ICTs.

Methods

Survey Design and Distribution

This survey study was conducted in the context of our media-health research, which focuses on designing personalized ICTs for coping with chronic mental and physical health problems. We invited participants to complete an anonymous online survey investigating the relation between screen addiction and health. The multifactorial survey included direct categorical self-assessment questions, as well as indirectly measured scales, to compare the estimated severity of problems (ie, health, screen usage, and stress) versus the individual's self-categorizations (stressed/not stressed; addicted/not addicted).

The minimum sample size of 355 was determined based on an expectation of 95% confidence level (5% margin of error) in receiving survey responses in a population of 2600—the size of subscribers to the PERFORM Centre's newsletters and email list of volunteers interested in studying the relation between lifestyle and health. The survey was provided in both French and English. We obtained institutional ethics approval for this study from Concordia University. All participants provided consent, and their participation was fully anonymous and with no remuneration.

The survey obtained demographic information (age, sex, ethnicity, years of education, and profession); Likert-scaled questions about the amount of usage of, dependence on, and importance of different screen types and related activities; and finally, questionnaires to assess vulnerability to different types of stress (details below).

Screen Variables

Screen Addiction

Participants were asked to report if they considered themselves addicted to *computer games or the internet*. If they responded *heavily* or *moderately*, they were categorized into self-admitting screen addicted referred to as *screen addicted (A1*), and if they responded *No*, they were categorized as *nonaddicted (A0)*. We also asked them to estimate the hours (less than 1 hour, 2 to 3 hours, 4 to 5 hours, more than 5 hours) they spent each day on screen-related leisure activities (television, internet, games, and watching videos on computer) to ensure the consistency of self-reported addiction and actual screen time. In addition, we administered a subset of Young's Internet Addiction test (IAT) [81] including the following items: (1) surfing the internet longer than you intended; (2) forgetting house chores while online; (3) loss of sleep due to internet activities; (4) more time spent online than with family; (5) work or grades suffering as a result of online activities; (6) defensive or secretiveness about being online; (7) nervousness and moodiness due to being offline; (8) preferring online activities over going out; (9) forming new relationships with fellow online users; and (10) others complaining about the amount of time spent online. Each question was scored on a 1 to 5 Likert scale (Never, Rarely, Sometimes, Frequently, or Always). Cronbach alpha on the selected IAT items was .869. The sum of the scores was used as a scale of *internet overuse.*

Screen Repertoires

In this report, *screen-repertoires* include electronic display surfaces on which visual content is projected or reflected (eg, a television set, a computer terminal, or a handheld electronic device, such as a tablet or a smartphone) and used for any of the following functions: generation or consumption of information, communication, or entertainment. To investigate the patterns of screen usage, we asked 3 sets of questions. To investigate *How* they use them (*Screen Importance*), they were asked to rate the importance of the following functions in their daily lives: (1) short message service (SMS) or email, (2) playing, (3) online shopping, (4) social networking, (5) searching for information, (6) following the news, (7) watching videos and movies, and (8) e-reading. To understand *Why* individuals use screens (*Screen Dependence*), they were asked to score their daily dependence on screens for the following needs: (1) education, (2) information, (3) entertainment, (4) relaxation, (5) social networking, (6) professional networking, and (7) work. Finally, to assess *What Technologies* they depend on (*Screen Necessity*), they were asked to indicate which technologies they needed to have access to on a weekend or during their vacation: (1) desktop, (2) laptop, (3) smartphone, (4) tablet, (5) e-reader, (6) television, and (7) game console. All questions were scored from 5-point Likert scales (strongly agree to strongly disagree) and were binarized to *High* for agree and strongly agree and *Low* for indifferent, disagree, and strongly disagree responses.

Stress Variables

Working Definition of Stress

Our working definition of stress draws from Mason's 1968 [82] and Dickerson and Kemeney's meta-analysis [83] that showed the perception of loss of control in presence of real or perceived self-threatening or unpredictable situations to be the common denominators of triggering a physiological stress reaction. The reason why we focus on this neurological definition is because we are interested in identifying technologies whose impact on stress can be empirically and quantitatively examined in the future. However, instead of focusing on a single stress questionnaire, we investigated 4 potential factors that are likely to be stressful: emotional stress (ES; presence of negative feelings), perceptual stress (anticipation of stressful loss of control and status in common life experiences), health stress (inability to perform normal daily functions), and life domain satisfaction (satisfaction with work, family, social support, finances, and leisure). Internal consistency of the questionnaires was established using reliability analysis. All scores were computed by summing up the Likert scores as described below. The final stress level was computed for each stressor as the percentage of the maximum possible score (ie, if someone expressed highest level of stress in responding to all questions). These ratio scores enabled us to conduct a relative comparison of different stressors' intensities.

Emotional Stress

Emotional stress refers to the state of a personal experience of negative mood and affect such as anxiety, anger, lack of motivation, sadness, or irritability. These mood states can be considered as internal risk factors that explain the interindividual vulnerability to stress. Individuals with mood and anxiety disorders are more stressed [84,85] and are at higher risk of negative health consequences as a result of chronic stress [16,86,87]. We estimated ES using a 5-item questionnaire, adapted from the Depression, Anxiety, and Stress Scale [69], asking participants to rate the following question: *During the past four weeks, how much have you been bothered by any* *emotional problems such as anxiety, sadness, lacking motivation, being sensitive and irritable, and anger* (scored on a Likert scale 0-3, Not at all; A little; Quite a bit; A lot). The Cronbach alpha value of standardized items was .86.

Perceptual Stress

Perceptual stress refers to the vulnerability to experiencing lack of control and perceiving a threat to ego while facing the external world. Unlike ES that measures the actual state of negative feelings and affect, perceptual stress reflects anticipation of a stressful experience. In one of the earliest meta-analysis studies of physiological manifestation of stress, Mason showed that the perception of novelty, unpredictability, lack of control, and threat to ego would reliably predict an autonomic and neuroendocrine response [82]. Dickerson and Kemeney's meta-analysis of 208 acute stress studies confirms that loss of control in time-limited cognitive tasks or public performance under social evaluative pressure is a reliable trigger of stress response [83]. A common questionnaire to measure perceptual stress is the perceived stress scale, which asks explicit questions about the individual's sense of control, irritability, uncertainty, and feeling stressed over the past month. We approached the question differently and aimed to assess the general degree of vulnerability to being stressed by commonly lived experiences. We aimed to assess interindividual variations in coping with unknown, unpredictable, and *ego*-threatening circumstances such as being in situations where one may lose control and be under time pressure (such as driving and working overtime) or be judged negatively, for example, in a job interview, public speaking, taking an exam, or going on a first date [82,83,88]. We constructed a 12-item questionnaire and asked the participants to rate how stressful they found the following general situations: (1) not having control (lack of control); (2) making decisions that affect you (uncertainty affecting self); (3) making decisions that affect others (uncertainty affecting others); (4) taking an exam (time/performance pressure), (5) being judged negatively (threat to ego); (6) giving a public speech (social evaluative threat); (7) driving (lack of control); (8) being overworked (lack of control); (9) being in a competition (time/performance pressure under social evaluative threat); (10) getting sick (lack of control); (11) going for a job interview (uncertainty affecting self and social evaluative threat); and (12) going on a first date (social evaluative threat). These items were each scored on a 4-point Likert scale (Not stressful at all, Not stressful, Somewhat stressful, Very stressful, and Extremely stressful). The Cronbach alpha value of standardized items was .81.

Health Stress

To be suffering from illness or chronic health conditions is a major stressor that is by and large outside an individual's locus of control. To evaluate whether individuals suffered health stress, we first asked them to rate the general state of their mental and physical health (good, bad, and could be better). We then asked 7 questions adapted from the Medicare Wellness Checkup survey [89] to assess whether they suffered conditions that would reduce their sense of control over normal daily functions of life. We asked participants to rate how often they have been bothered by any of the following conditions: (1) falling or feeling dizzy when standing up; (2) sexual problems; (3) trouble eating well; (4) problem using the telephone; (5) problem using the computer; (6) problem driving; (7) problem reading; and (8) tiredness or fatigue (scored on a Likert scale 0-4: Never, Rarely, Occasionally, Often, Always). The Cronbach alpha value based on standardized items was .729.

Life Dissatisfaction

External factors such as family, friends, work conditions, and financial situations are important well-being factors [90] that can moderate the severity of stress. Supportive and satisfactory personal and professional networks can mitigate adverse effects of health or ES. In contrast, financial, professional, and relationship problems (at work and at home) which are outside an individual's perceptual, emotional, or practical control can burden their ability to maintain control over their own life. We considered work, family, social relationships, financial comfort, and leisure as life domains to have a potentially significant impact on stress levels. A 6-item questionnaire asked participants to rate how satisfied they felt with the following: (1) My boss is friendly and fair; (2) My work and leisure activities are balanced; (3) My family is supportive; (4) My friends are there for me; (5) My life is under control financially; (6) My work and/or studies are enjoyable (scored on a Likert scale of 0-3, ranging from *true* to *not true at all*). The Cronbach alpha value of standardized items was .72.

Self-Reported Stress

To examine the correspondence between our stress variables and the individual's subjective evaluation of stress, we asked them to report whether they have experienced a recent stressful event. Self-reported stress was a binary variable based on a *Yes/No* (hereafter referred to as *stressed/not stressed*) response to *Have you experienced a major stressful episode in the past year?* We then refined the question by asking participants to check which type of stress they had suffered: bereavement, financial hardship, job loss, school exams, chronic health problems, and relationship problems. These stressors, in terms of their psychophysiological impact, are not equivalent; however, we wanted to capture the heterogeneity in the subjective perception of stress and to compare the intensity of our quantitative metrics in relation to these major life stressors.

Statistical Analysis

Statistical analyses and visualizations were performed with SPSS 24.0 (IBM, SPSS Statistics, for OX) and Prism 7.0 (Graphpad Inc, for OX).

Univariate statistics were presented as percentages of the response frequencies. Two-way analysis of variance (ANOVA) was used to assess the effect of subjective addiction and stress interaction on screen usage and stress scores. The Kruskal-Wallis test was performed to compare group differences in rating the screen importance, dependence, and access. A posthoc *t* test was used (with Welch-Satterthwaite correction to adjust degrees of freedom for cases where equality of variance was violated) to compare differences in stress scores in relation to rating dependence on different screen-related activities: high or low. Finally, to examine the best model that explained the likelihood of belonging to the screen addicted group, we performed a logistic regression (including different stress scores, age, and gender as explanatory variables). We also tested the same model factors in a regression model with internet overuse as a dependent variable. Statistical significance was set at .05. A test of collinearity was performed to ensure that the variance inflation factor (VIF) was below 3.

Results

Sample Characteristics

Out of 654 responders, the final sample size (based on complete case of all variables of interest, that is, screen addiction and stress scores) was 459. Sample characteristics are presented in Table 1. See Multimedia Appendix 1 for more details on the intensity of stress in each category.

Group Differences in Stress and Screen Usage Scores

Figure 1 (A) illustrates the overlaps in subjective evaluations of stress and screen addiction. Approximately, 30% of the sample considered themselves screen addicted. The majority of the sample reported recent stress but no addiction S1A0 (44.6%, 205/459), followed by S0A0 (25.9%, 119/459), S1A1 (19.8%, 91/459), and S0A1 (9.5%, 44/459). Figure 1 (B) summarizes the ANOVA results. There was no stress by addiction interaction effect on any of the variables ($F_{1,455}$<3, P>.1). Individuals who reported themselves as stressed differed in ($F_{1,455}$=5.98, P=.02 emotional ($F_{1,455}$=25.4, P<.001), perceptual ($F_{1,455}$=9.49, P=.002), health ($F_{1,455}$=11.7, P<.001), and life dissatisfaction ($F_{1,455}$=13, P<.001), internet overuse ($F_{1,455}$=6.83, P<.01) but not screen time. In terms of self-rated screen addiction, groups differed in age ($F_{1,455}$=54.3, P<.001), ES ($F_{1,455}$=40.4, P<.001), perceptual ($F_{1,455}$=11, P<.001), health ($F_{1.455}$=23.1, P<.001), and life dissatisfaction ($F_{1,455}$=29.7, P<.001), internet overuse ($F_{1,455}$=142, P<.001), and screen time ($F_{1,455}$=70.2, P<.001). Only 23% of variation in internet overuse and 13% of variation in screen time was explained by self-admitted addiction. With the exception of screen time (where no difference between the stressed and nonstressed was observed), the S0A0 group (ie, those who were not stressed and not addicted) reported significantly lower stress and screen usage compared with S1A1 (those who were both addicted and stressed; more details in Multimedia Appendix 2).

Screen Repertoires With Respect to Self-Reported Addiction and Stress

Figure 2 illustrates group differences in the salience of screens in daily life in terms of the frequency of rating a screen or screen-related activity *high*. Results of the Kruskal-Wallis test are presented in Table 2. Significant group differences emerged in rating the importance of daily usage of social networks (highest in S1A1), games (highest in S0A1), and e-books (highest in S0A1). Groups differed in rating the necessity of access to desktop computers (highest in S1A1), laptops (highest in S0, regardless of A), mobile phones (S1A1), smartphones (highest in A1, regardless of S), and game consoles (highest in A1, regardless of S). Groups differed in daily dependency on screens for education (highest in A1, regardless of S), entertainment and relaxation (highest in S0A1), social communications (highest in A1, regardless of S), and professional networking (highest in S1A0).

Table 1. Sample characteristics.

Variable	Statistics
Age (years), mean (SD)	36 (14)
Younger than 36, n (%)	272 (60)
Sex, n (%)	
Male	136 (30)
Female	323 (70)
Education, n (%)	
CEGEP	102 (23)
Bachelor’s	201(44)
Post graduate	150 (33)
Ethnicity, n (%)	
White	344 (74)
Hispanic	15 (3)
Black	22 (5)
Asian	27 (6)
Middle Eastern	34 (7)
Other	16 (4)
Profession, n (%)	
Student	122 (26)
Educator	58 (13)
Office professional	50 (11)
Wellness and health care	48 (11)
Artist	46 (10)
Other	135 (29)
Stressed, n (%)	
No	185 (40)
Bereavement	64 (14)
Financial	120 (26)
Exam	107 (23)
Chronic health	69 (15)
Relationship	135 (29)
Mental health, n (%)	
Good	272 (60)
Bad	14 (3)
Could be better	170 (37)
Physical health, n (%)	
Good	277 (61)
Bad	15 (3)
Could be better	167 (36)

Figure 1. (A) Distribution of the sample based on self-rated screen addiction and recent stress; (B) Group differences in age; (C) Group differences in internet overuse; (D) Group differences in daily screen time; (E) Group differences in emotional stress; (F) Group differences in perceptual stress; (G) Group differences in health stress; (H) Group differences in life dissatisfaction; (I) Posthoc estimated mean differences of stress based on Screen x Addiction categories. The largest differences are observed in comparison of S0A0 (neither addicted nor stressed) versus S1A1 (both addicted and stressed).

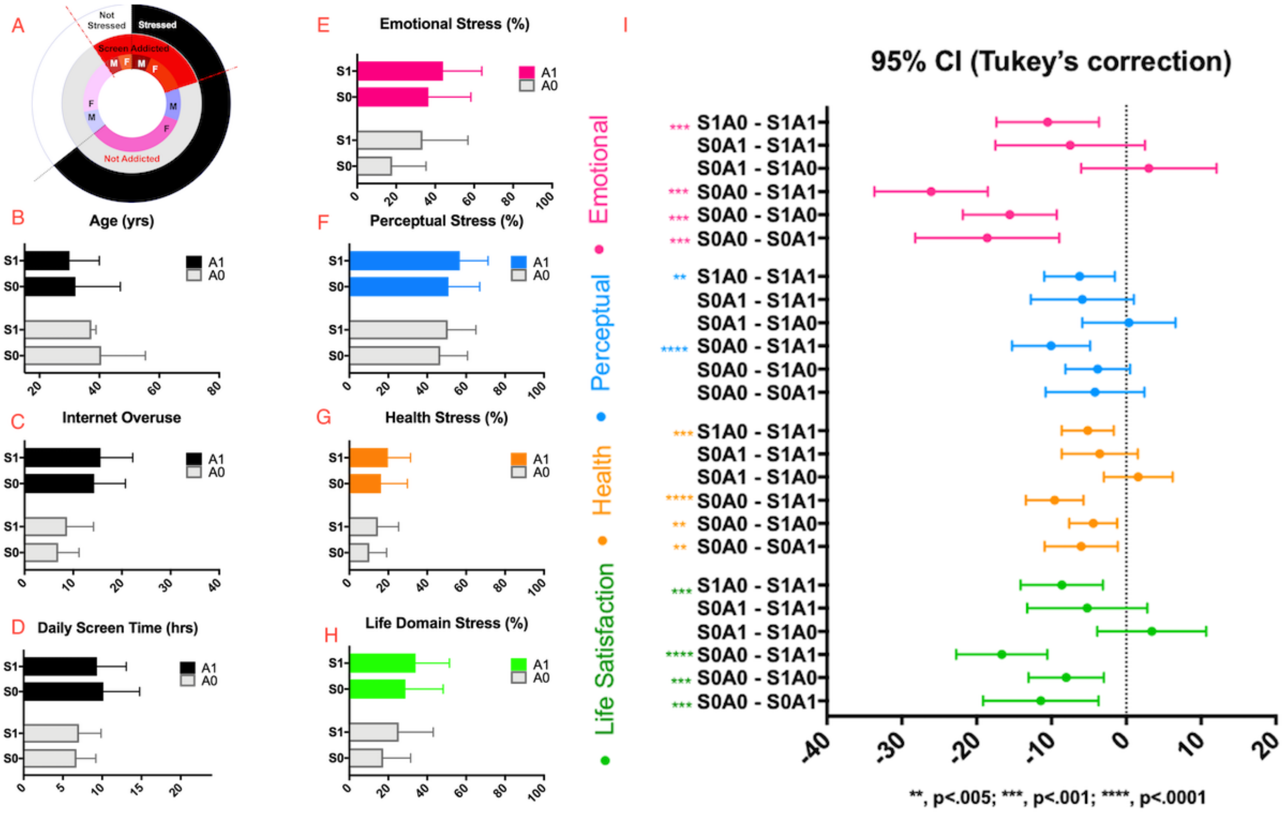


Figure 2. Radar diagram of the ratio of individuals within each group who rated various aspects of the screen repertoires highly; (A) How important are these activities to your daily life? (B) What is the degree of daily dependence on screens for these activities? (C) How essential is it to access these screens on a holiday or during the weekend.

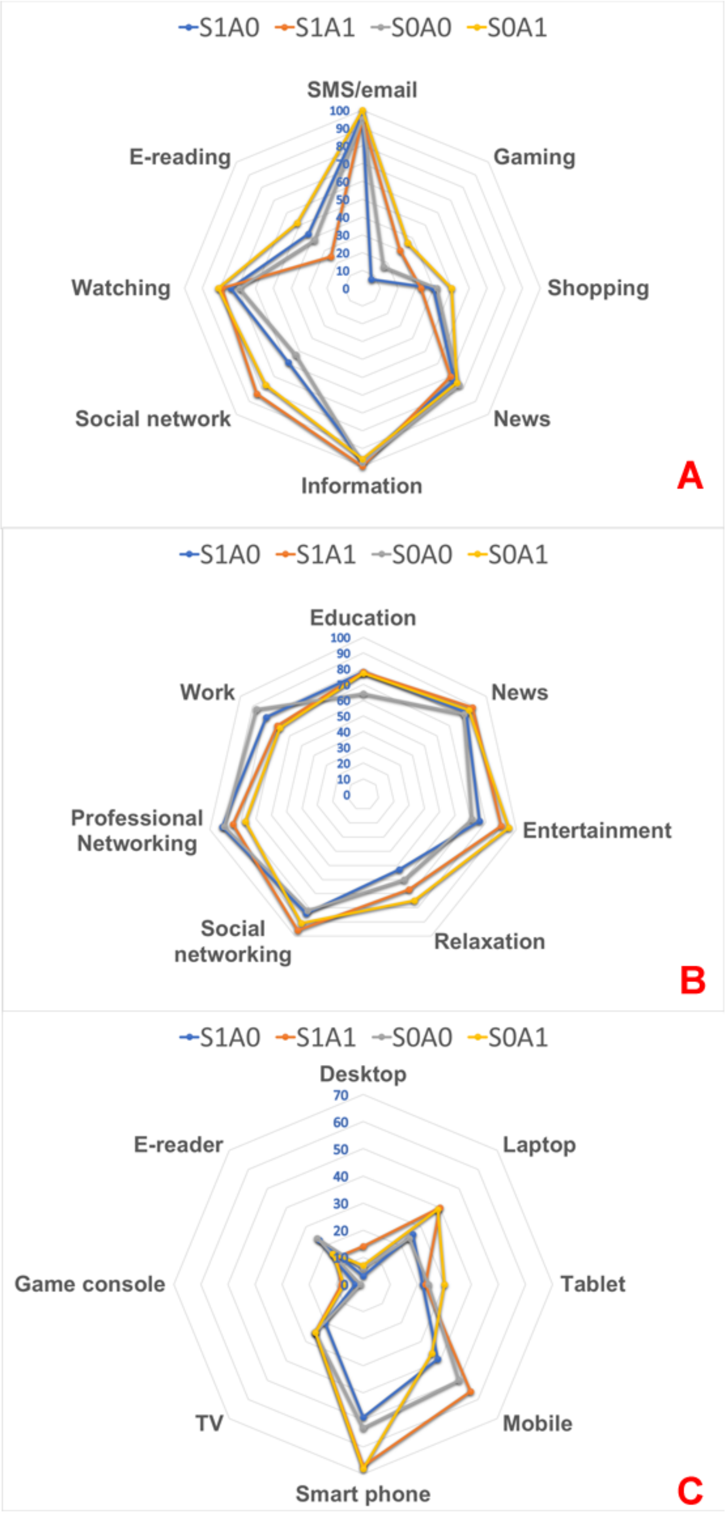


Table 2. Results of the Kruskal-Wallis test. Percentages of individuals in each category rating each item high or very high (N_{High}) are listed (S0, no recent stress; S1, reported recent stress; A0, not addicted; A1, screen-addicted [self-rating]).

Screen-related ratings per groups	S0A1 N_{High}/44 (%)	S0A0 N_{High}/119 (%)	S1A1 N_{High}/91 (%)	S1A0 N_{High}/205 (%)	X^2, *df*=3	*P* value
Screen dependence						
Education	34 (77.3)	76 (63.9)	71 (78.0)	160 (78.0)	9.1	.03
News and info	38 (86.4)	98 (82.4)	81 (89.0)	173 (83.9)	3.85	.27
Entertainment	42 (95.5)	84 (70.6)	82 (90.1)	155 (75.6)	20.5	<.001
Relaxation	33 (75.0)	71 (59.6)	61 (67.0)	109 (53.2)	9.9	.02
Social networking	40 (90.9)	97 (81.5)	87 (95.6)	172 (83.9)	10.7	.01
Professional communications	34 (77.3)	108 (90.8)	77 (84.6)	188 (91.7)	9.6	.02
Work	30 (68.2)	103 (86.6)	64 (70.3)	163 (79.5)	11.02	.01
Screen importance						
Short message service (SMS) or email	44 (100)	113 (95.0)	86 (94.5)	198 (96.6)	2.9	.4
Game	16 (36.4)	20 (16.8)	27 (29.7)	15 (7.3)	35.6	<.001
Shopping	22 (50.0)	50 (42)	30 (33.0)	83 (40.5)	3.89	.27
Following the news	33 (75.0)	92 (77.3)	64 (70.3)	150 (73.2)	1.4	.71
Searching	42 (95.5)	117 (98.3)	91 (100)	199 (97.1)	3.85	.28
Social media	34 (77.3)	63 (52.9)	76 (83.5)	122 (59.5)	26.5	<.001
Watching	36 (81.8)	82 (68.9)	72 (79.1)	152 (74.1)	4.23	.24
e-reading	23 (52.3)	45 (37.8)	23 (25.3)	88 (42.9)	11.8	.008
Screen necessity						
Desktop computer	3 (7.1)	7 (5.9)	13 (14.3)	7 (3.5)	12.21	.007
Laptop	17 (38.6)	28 (23.5)	36 (39.6)	52 (25.5)	9.83	.02
Tablet	13 (30.2)	28 (23.5)	21 (23.3)	46 (22.4)	1.2	.75
Smartphone	30 (68.2)	62 (53.0)	60 (66.7)	99 (48.8)	11.39	.01
Mobile phone	16 (36.4)	60 (50.4)	50 (55.6)	80 (39.4)	9.32	.03
Television	11 (25.0)	29 (24.6)	23 (25.3)	40 (19.5)	1.9	.59
Game console	3 (6.8)	1 (0.8)	7 (7.7)	5 (2.5)	9.32	.03
e-reader	7 (15.9)	29 (24.4)	13 (14.4)	46 (22.5)	4.1	.25

Different Stressors and Different Screen Dependencies

Differences in various stress types and screen overuse were examined based on dependence (low/high) on different screen-related activities. With the exception of professional networking, all other screen activities were associated with higher internet use. Screen time was not different in relation to depending on screens for professional networking or work (Table 3).

Those who highly depended on screens for entertainment and relaxation had significantly greater levels of perceptual stress. The ES was higher in those who depended on screens for entertainment, social networking, and education. In addition to ES, those who depended on screens for social networking also had higher levels of perceptual and health stress. In contrast, those who depended on screens for professional networking had lower scores of life dissatisfaction and no differences in other stressors. In those who depended on screens for work, both life dissatisfaction and perceptual stress scores were lower (Table 4).

Heterogeneity of Stress Types and Screen Dependencies

Finally, to examine the heterogeneity in subjective perception of stress and how that would relate to variations in screen dependency, we compared groups based on the type of stress that they reported (Table 5). With the exception of bereavement, irrespective of stress type, all stress scores were higher in those who reported suffering stress in the past year. With the exception of those reporting bereavement and chronic health problems, those who reported other types of stress were younger and had greater scores of internet overuse. Differences in stress levels are presented in Multimedia Appendix 1.

Table 3. Differences in screen usage based on screen dependence (number of individuals rating screen usage as high [N_{high}] vs number of individuals rating screen usage as low [N_{low}]).

Dependence	Screen usage	Internet use		Screen time	
	N_{high}/N_{low}	95% CI	*P* value	95% CI	*P* value
Education	344/18	0.1 to 2.8	.03	0.3 to 1.7	.004
News	389/70	.02 to 3.3	.05	0.1 to 1.8	.03
Entertainment	363/96	2.3 to 5.2	<.001	0.3 to 1.8	.008
Relaxation	274/185	1.5 to 3.8	<.001	0.1 to 1.33	.03
Social networking	396/63	2.6 to 5.9	<.001	0.8 to 2.54	<.001
Professional networking	407/52	—[a]	—	—	—
Work	360/99	−2.0 to 0.72	.005	—	—

[a] No significant difference was observed.

Table 4. Differences in stress levels based on screen dependence (number of individuals rating screen usage as high [N_{high}] vs number of individuals rating screen usage as low [N_{low}]).

Dependence	Screen usage	Emotional		Perceptual		Illness		Life dissatisfaction	
	N_{high}/N_{low}	95% CI	*P* value	95% CI	*P* value	95% CI	*P* value	95% CI	*P* value
Education	344/18	3.4 to 12.9	<.001	—[a]	—	—	—	0.5 to 5.1	.01
News	389/70	—	—	—	—	—	—	—	—
Entertainment	363/96	1 to 11.3	.01	0.7 to 7.4	.01	—	—	—	—
Relaxation	274/185	—	—	3.6 to 9.1	<.001	—	—	—	—
Social networking	396/63	7.5 to 19.5	<.001	7 to 14.8	<.001	—	—	1.8 to 7.7	.005
Professional networking	407/52	—	—	—	—	−17.5 to −5.4	<.001	—	—
Work	360/99	—	—	−8.2 to −1.5	.005	−9.9 to −0.6	.03	—	—

[a] No significant difference was observed.

Table 5. Differences between stressed and nonstressed groups based on the type of stress. Values represent the estimated mean differences between those reporting the type of stress versus others. Stress scores are normalized to the highest possible score for each stress type.

Stress type	Age	Internet overuse	Screen time	Emotion	Perception	Health	Dissatisfaction
Bereavement	—[a]	—	—	—	—	—	—
Financial	−5.8[b]	2.6[b]	—	13.9[b]	4.6[c]	5.14[d]	13.7[d]
Exam	−14.9[b]	3.9[b]	1.06[c]	19.9[b]	10.7[d]	7.4[d]	8.4[d]
Chronic health	—	—	—	9.6[d]	5.4[c]	5.5[d]	8.5[c]
Relationship	−4.17[c]	1.7[c]	0.8[c]	12.2[b]	3.1[c]	3.6[c]	7.5[d]
Mental health	−7.4[d]	3.8[d]	0.9[c]	29.8[b]	10.5[b]	9.7[b]	15.3[b]
Physical health	−2.9[c]	3.4[b]	—	15.3[b]	6.5[b]	8.6[b]	11.2[b]

[a] No significant difference was observed.

[b] $P<.001$.

[c] $P<.05$.

[d] $P<.01$.

Table 6. Parameter estimates for logistic regression, dependent variable: screen addiction.

Model A	Regression coefficients (B)	Standard error	Odds ratio (95% CI)	*P* value
Age	−.039	0.010	0.962 (0.944 to 0.980)	<.001
Emotional stress	.005	0.006	1.005 (0.993 to 1.017)	.43
Perceptual stress	.015	0.009	1.015 (0.997 to 1.032)	.09
Health stress	.022	0.012	1.022 (0.999 to 1.045)	.07
Life dissatisfaction	−.023	0.006	0.978 (0.967 to 0.989)	<.001
Sex (F<M)	.898	0.258	2.454 (1.47 to 4.07)	.001

Table 7. Parameter estimates for linear regression, dependent variable: internet overuse.

Model B	Standardized coefficients (B)	Standard error	*t* value	95% CI for B	*P* value
Age	−0.223	0.019	−5.28	−0.137 to −0.063	<.001
Emotional stress	0.165	0.015	3.14	0.017 to 0.075	.002
Perceptual stress	0.139	0.020	3.04	0.021 to 0.098	.002
Health stress	0.111	0.027	2.34	0.010 to 0.117	.02
Life dissatisfaction	0.200	0.013	4.62	0.035 to 0.087	<.001
Sex (F<M)	−0.147	0.581	−3.56	−3.2 to −0.913	<.001

We found significant differences in dependence on social networks in individuals who reported financial, exam relationship, and mental health stresses compared with the nonstressed, those who reported exam stress were more dependent on screens for education, entertainment, relaxation, and social networking but less for work. To control for the sample bias, we reran the analyses on 188 samples, after excluding all students. In this case, the only significant difference in screen dependency was in social networking (higher in those with financial stress; $t_{186}=2.7$, $P<.01$). In this subgroup, age, internet usage, and screen time were not dependent on stress, but robust differences in stress scores (ie, those with $P<.001$) were also observed in this subgroup.

Regression Analysis of Age, Gender, and Stress in Relation to Subjective Screen Addiction and Quantitative Internet Overuse

Having identified group heterogeneities both in terms of different stressors and different screen-related activities, we investigated to what extent variables such as age, gender, and various stress factors explained the likelihood of being screen addicted (logistic regression, Model A, see Table 6) and internet overuse (linear regression, Model B, see Table 7). The VIF for independent variables was below 1.5, thus the model had sufficient tolerance to collinearity. Cross-correlation coefficients are provided in Multimedia Appendix 2.

In Model A (which explained 24% of the variance in the likelihood of being in the A1 group), to be younger, male, and have lower life satisfaction were the most important predictors of the likelihood of identifying oneself as screen addicted. In Model B (which explained 30% of variations in internet overuse), all factors were significant, with age being the strongest factor followed by dissatisfaction, ES, gender, perceptual stress, and finally health stress.

Discussion

Principal Findings

We examined the relation between screens and stress using the repertoire-oriented media research framework and showed significant associations between self-admitted screen addiction and quantitative stress levels, as well as stress-specific usage of screens. Individuals who consider themselves screen addicted are also more stressed and are more likely to use screens for entertainment and social networking.

An important finding is that the relation between subjective and objective self-assessment of stress and screen addiction is not overlapping. Although 65% of survey responders reported having suffered various stressful events, only one-third of those also considered themselves *screen addicted.* Self-reported stress did not predict significant likelihood of belonging to the screen-addicted group either. However, certain stressors such as financial, relationship, exam, and health problems were associated with higher rates of screen overuse. Interestingly, the magnitude of the estimated difference in stress levels was larger in the self-admitting addicted versus the nonaddicted group, compared with the self-admitting stressed versus nonstressed group, suggesting an implicit link between actual stress and perception of screen addiction. The screen addicted group had significantly higher internet use and screen-time scores, therefore confirming that subjective assessment of screen addiction corresponded to actual usage metrics, but the average scores were not very high; therefore, it is unlikely that any of the participants were problematic screen users.

In addition, we explored differences in the pattern of screen repertoires in different subgroups. The general patterns were similar between all groups, with following the news and gathering information being the highest and equally important activity in all groups. The necessity of access to smartphones was the highest, and the necessity of access to game consoles, followed by television, the lowest. The pattern of daily dependence on various functions was equipotent (above 70%) across all possible activities, but the center of the pattern shifted toward entertainment and relaxation for the self-admitting addicted and toward work and professional networking for the nonaddicted. The strongest pattern differences emerged at the level of the importance of social networking and gaming and dependence on screens for entertainment. In fact, emotional, perceptual, and health stress were significantly higher in those who depended on screens for social networking. Perceptual stress was also higher in individuals who used screens for entertainment and relaxation. In contrast, individuals who used the screens for work had lower perceptual stress and higher life satisfaction. Post hoc analysis of the effect of subtypes of stress on screen dependency further confirmed that social networking was important to those reporting financial, relationship, mental health, and exam stress. The fact that exam stress was the only type of stress to predict differences in dependence on entertainment and relaxation indicates a demographically specific effect exclusive to a younger student subsample. However, after excluding students from the sample, differences in social networking related to financial stress were still significant.

Our regression analyses show that age and gender influence the prevalence of screen dependency, but only a small portion of variations in screen addiction (24%) or internet overuse (30%) was explained by stress and demographic factors; therefore, other variables must contribute to individual’s screen usage patterns.

Comparison With Previous Work

Robust Association Between Stress and Social Networking

In our sample, over 90% of screen addicted and over 80% of the nonaddicted considered social networking highly important, consistent with global statistics, indicating that the percentage of adults using social media has reached 94% [91,92]. Griffith and Szabo have shown that social networking is the most prevalent of all online activities [65]. In our study, variations in social networking emerged as the most robust indicator of reporting oneself as screen addicted, concurrent with having high levels of emotional and perceptual stress—mainly in the young students and also in a subgroup who reported financial stress. In general, the proportion of self-admitting, stressed individuals who rated social networking important was high. Individuals who depended on screens for social networking had larger scores of emotional, perceptual, and life dissatisfaction. These findings corroborate an earlier review of uses and gratifications research that revealed that individual’s dependence on social media related to their need for relationship maintenance, passing time, entertainment, and companionship [93]. It has been argued that through a myriad of stimulation and interaction, social media can modify mood for better or for worse [38,94,95]. Could social networking have caused the higher stress levels? It has been shown that spending time on Facebook causes a decrease in mood by increasing envy and reducing the social capital [48,49], increasing anxiety about relationships [96,97], or increasing guilt about having wasted time [98]. Our cross-sectional study design precludes any conclusions about the causality of the relationships we report, but it is plausible to suggest that those who consider themselves screen addicted perceive time spent on social networks with a more negative connotation, than time spent on following the news or searching information—which were both the most important activities for all groups.

Habitual Screen Use for Coping With Stress

We observed a relative shift toward depending on screens for entertainment, relaxation, social networking, and education in the subjectively stressed and screen addicted group (S1A1). This group had approximately 26% greater ES compared with the S0A0. In contrast, dependency on screens in the S0A0 shifted toward work and professional networking. When asked about the importance of a set of activities on a regular basis, in addition to social networking, playing games was important to 36% of S0A1 group, and 31% of S1A1, versus 17% of the S0A0, suggesting that gaming was not related to the experience of stress. Nevertheless, approximately 8% of the S1A1 group considered access to game consoles necessary, as opposed to only 1% of the S0A0. One possible explanation is that using games serves as a coping strategy against anxiety, which is consistent with the observation of significantly higher associations between emotional and perceptual stress and greater dependency on screens for entertainment and relaxation (particularly in students). This interpretation is in line with previously reported comorbidities between anxiety and depression and excessive use of games or internet, mainly in the young [58-64,99,100] This raises the question, should the excessive usage of screens (games and social networks) for relaxation and entertainment be considered as addiction? Is this cultural connotation the reason why those who use screens for leisure activities are more likely to label themselves as screen addicted? Or is it because spending too much time on screens distracts students from the school work and thus becomes a stressor?

There is controversy whether behavioral compulsions should be treated as addiction disorders or as an individual’s adaptive choices [73,101] that should be dealt with by accounting for socio-relational heterogeneities [72]. Weisel et al have previously suggested that screen *addiction* is not necessarily a problematic phenomenon (as it is commonly referred to in the literature), but a manifestation of the individual’s coping strategy, which should be channeled toward care [24]. Despite the evidence that social networks such as Facebook can be stressful [48,49,96-98], there exists some empirical evidence to suggest that being connected to social media can mitigate the physiological response to a psychosocially stressful condition [38] or that adding social media interventions improves the outcome of traditional psychiatric treatment of depression [39]. Games also interact with myriad cognitive, executive, and rewards processes, and as such they have a quantifiable impact on physiological stress response [102,103]. That perceptual stress (measured by questions about the degree of feeling stressed by control and ego-threatening situations) was higher in those who relied on screens for entertainment and relaxation is noteworthy. Recall that we defined stress as the body’s adaptive response to restore mental or physiological balance while challenged by external or internal, perceived or real, threats to self [88]. This biological framework is important in interpretation of screen-related stress, because the inverted U-shape of stress response determines which kind of activity causes or diminishes the physiological response. In dealing with stress, individuals adopt problem-focused or emotion-focused coping styles, that can be based on either avoidance or approach to confront a stressor [13,18]. To play games for mental disengagement following a major life event may be a form of problem-focused coping [20].

Should we rethink the generally negative connotation in linking stress to screen-*addiction*? In our sample, approximately 70% the S0A0 depended on screens for entertainment, and the highest ratio of dependence on screen for entertainment (95%) was S0A1, that is, those who considered themselves screen addicted but not stressed. Recall that despite the fact that stress levels were high in screen addicted individuals, stress, age, and gender explained no more than 30% of variations in internet overuse and no more than 24% of the likelihood of being in the group who identified themselves as screen addicted. Also, recall that to use screens for work, information and news were the most salient of activities for everyone, but they did not differ across groups. Therefore, it is perhaps not the amount of screen usage, but subjective differences in justification of using screens for leisure activities that explain our results. Future studies are needed to explore these questions in relation to personality and perceptual factors.

New Contributions

Subjective Rating of Screen Addiction in Relation to Different Types of Stress

To the best of our knowledge, this is the first study to examine the relation between subjective assessment of screen addiction, various sources of stress, and various screen-related activities. Approximately, one-third of our sample identified themselves as screen addicted, and indeed the scores of internet overuse and screen time in this group were significantly greater than the nonaddicted. Although it should be noted that the scores were not near the maximum, suggesting that the self-assessments reflected a personal perception of inadequate screen usage, rather than an actual *abuse*. To have accounted for both subjective and quantitative measures of stress and addiction revealed interesting differences in the magnitude of stress based on stress category versus addiction category. Interestingly, self-admitted screen addiction revealed greater difference in all stress categories, compared with reporting recent stress. To account for this subjective difference is important particularly in the context of studying the relation between stress and screen use.

We took a similar multivariate approach to stress as well and found that serious stressors such as bereavement and chronic health were not associated with differences in internet overuse, screen time, or any daily screen dependency, but financial, exam, relationship, and self-evaluated mental health stressors were associated with greater dependency on social networking. This multifactorial approach helped identify subtle differences in the type of stress-screen association. For instance, we found that dependence on screens for entertainment and social networking was associated with greater emotional and perceptual stress in contrast to dependence on screens for work, which was associated with lower scores of dissatisfaction and smaller perceptual stress scores. Although different stress scores shared some variance (no more than 28%), they were not strongly collinear, and each measured different sources of vulnerability. Emotional stress reflected the impact of existing and prospective feelings and anxieties that are experienced by an individual. Perceptual stress coded interindividual variations in self-confidence to cope with unknown, unpredictable, and *ego*-threatening circumstances such as being in situations where one may be judged negatively, that is, in a job interview, public speaking, taking an exam, or going on a first date [82,83,88]. Life satisfaction related to external factors that include relationships at work and with friends and family, financial control, and work/leisure satisfaction. These findings underline the necessity of designing experiments that account for personal and social variants that account for population heterogeneities in media selectivity, resilience, and coping [72-74].

Patterns of Screen Usage in Relation to Different Types of Stress

To the best of our knowledge, this is the first study to have applied the media-repertoire framework and to have studied the interrelations between various screens in relation to stress. Griffiths et al have argued that in studying behavioral dependence on screens, the heterogeneity of activities ranging from news and shopping to gaming, social networking, etc must be accounted for [104]. To evaluate this view, we examined the prevalence of various types of screens (such as TVs, computers, and e-readers and tablets) or screen-related activities such as watching videos and movies, reading, and working, which provide a comparative reference to guide designing better screen technologies and interfaces for stress management. In our sample, we did not observe any group difference in using tablets, TV, or e-readers; nor in activities such as email or SMS, following the news, watching videos, or reading online. The dependence on screen for education or gathering information and news was also not different. Instead, the differences were significant in the importance of gaming and social networking, in dependence on screens for relaxation, entertainment, and social networking, and in the necessity of access to game consoles, mobile and smartphones, and computers on a weekend. These findings suggest that portability, communication, and leisure are important features for those with higher levels of emotional and perceptual stress, thus confirming the potential of digital health mobile apps for mitigating stress through ICTs [105].

Limitations

Sampling biases in this study confound interpretations. We designed this study to target internet-literate and self-conscious individuals who are concerned about the negative impact of screen-addiction on health. Our snowball survey method produced an age- and sex-biased sample, in which the majority of responders were female, younger than 36 years of age, white, and university-educated. This biased sampling is common in digital surveys and reveals which demographics are more likely to utilize and benefit from ICT in health intervention. Although it limits the generalizability of the conclusions in designing global digital health solutions, it also underlines the potential to work toward creating more inclusive digital ecologies.

These findings should be considered as an exploratory approach to the investigation of the interactions between screens and stress—both highly relevant for public health innovations. However, our data warrant no clinical interpretations. The term *screen addiction* must be interpreted exclusively in the context of subjective self-evaluation. Although we showed that self-admitting screen addicts have significantly greater screen time and internet usage, our survey is limited in revealing dimensions of salience, tolerance, and emotional dependence to screens. It is also limited in explaining the relation between screen usage and health states, which would be necessary for clinical categorization of a behavioral addiction.

Finally, no inferences about the causality of the relationships between stress and screen use are justified. Instead, we emphasize the heterogeneity of stress-related factors that can moderate screen-related behaviors. Our findings underline the importance of multivariate examination of screen dependency within various psychological or sociological context. Future studies are needed to explore socioeconomic and intergenerational variations more closely.

Conclusions

As McLuhan predicted, electronic media (television, at the time) is *an extension of our physical bodies*, which interacts with our adaptation system to restore our physical and psychological equilibriums. Our interdisciplinary approach provided evidence for the contextual heterogeneity of the relation between screens and their role in stress adaptation, specifically via online activities for entertainment and social networking. Future work needs to examine the clinical implications of these findings and explore the mediating effects of a screen-related lifestyle on mental health outcomes.

Acknowledgments

The funding for this study is provided through a fellowship from the Office of Research of Concordia University to NKM. NKM is jointly supported by McGill Centre for Integrative Neuroscience and PERFORM Centre, Concordia University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Differences in stress levels and screen addiction with respect to the type of major stress suffered in the previous year.

[PDF File (Adobe PDF File), 163KB - jmir_v21i4e11485_app1.pdf]

Multimedia Appendix 2

Tests of correlation and collinearity between different stress and screen variables.

[PDF File (Adobe PDF File), 183KB - jmir_v21i4e11485_app2.pdf]

References

1. McLuhan M. Understanding Media: The Extensions of Man. Canada: McGraw-Hill; 1964.
2. Selye H. The Stress of Life. New York, US: McGraw-Hill; 1956.
3. Selye H. Stress and the general adaptation syndrome. Br Med J 1950 Jun 17;1(4667):1383-1392 [FREE Full text] [doi: 10.1136/bmj.1.4667.1383] [Medline: 15426759]
4. Juster R, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. Neurosci Biobehav Rev 2010 Sep;35(1):2-16. [doi: 10.1016/j.neubiorev.2009.10.002] [Medline: 19822172]
5. Kim JI, Yun J, Park H, Park S, Ahn Y, Lee H, et al. A mobile videoconference-based intervention on stress reduction and resilience enhancement in employees: randomized controlled trial. J Med Internet Res 2018 Oct 22;20(10):e10760 [FREE Full text] [doi: 10.2196/10760] [Medline: 30348630]
6. Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, et al. Effectiveness of an internet- and app-based intervention for college students with elevated stress: randomized controlled trial. J Med Internet Res 2018 Apr 23;20(4):e136 [FREE Full text] [doi: 10.2196/jmir.9293] [Medline: 29685870]
7. Zhu B, Hedman A, Feng S, Li H, Osika W. Designing, prototyping and evaluating digital mindfulness applications: a case study of mindful breathing for stress reduction. J Med Internet Res 2017 Jun 14;19(6):e197 [FREE Full text] [doi: 10.2196/jmir.6955] [Medline: 28615157]
8. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]

9. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41. [doi: 10.2196/jmir.6397] [Medline: 28193600]
10. Chang F, Chiu C, Chen P, Miao N, Chiang J, Chuang H. Computer/mobile device screen time of children and their eye care behavior: the roles of risk perception and parenting. Cyberpsychol Behav Soc Netw 2018 Mar;21(3):179-186. [doi: 10.1089/cyber.2017.0324] [Medline: 29293374]
11. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
12. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
13. Lazarus RS. Emotions and interpersonal relationships: toward a person-centered conceptualization of emotions and coping. J Pers 2006 Feb;74(1):9-46. [doi: 10.1111/j.1467-6494.2005.00368.x] [Medline: 16451225]
14. Lazarus RS. Coping theory and research: past, present, and future. Psychosom Med 1993;55(3):234-247. [doi: 10.1097/00006842-199305000-00002] [Medline: 8346332]
15. Folkman S, Lazarus RS, Dunkel-Schetter C, DeLongis A, Gruen RJ. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. J Pers Soc Psychol 1986 May;50(5):992-1003. [doi: 10.1037/0022-3514.50.5.992] [Medline: 3712234]
16. Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat Rev Neurosci 2009 Jun;10(6):434-445. [doi: 10.1038/nrn2639] [Medline: 19401723]
17. Roth S, Cohen LJ. Approach, avoidance, and coping with stress. Am Psychol 1986 Jul;41(7):813-819. [doi: 10.1037/0003-066X.41.7.813] [Medline: 3740641]
18. Folkman S, Lazarus RS. The relationship between coping and emotion: implications for theory and research. Soc Sci Med 1988;26(3):309-317. [doi: 10.1016/0277-9536(88)90395-4] [Medline: 3279520]
19. Seery MD, Leo RJ, Lupien SP, Kondrak CL, Almonte JL. An upside to adversity?: moderate cumulative lifetime adversity is associated with resilient responses in the face of controlled stressors. Psychol Sci 2013 Jul 1;24(7):1181-1189. [doi: 10.1177/0956797612469210] [Medline: 23673992]
20. van Ingen E, Utz S, Toepoel V. Online coping after negative life events. Social Science Computer Review 2016 Aug 3;34(5):511-529. [doi: 10.1177/0894439315600322]
21. Petko D, Egger N, Schmitz FM, Totter A, Hermann T, Guttormsen S. Coping through blogging: a review of studies on the potential benefits of weblogs for stress reduction. Cyberpsychology (Brno) 2015 Aug;9(2):- [FREE Full text] [doi: 10.5817/Cp2015-2-5]
22. Saleh D, Camart N, Romo L. [Stress management intervention via the internet among students: review of the literature]. Ann Méd Psychol 2017 Apr;175(4):327-331. [doi: 10.1016/j.amp.2016.03.013]
23. Ryan C, Bergin M, Chalder T, Wells J. Web-based interventions for the management of stress in the workplace: focus, form, and efficacy. J Occup Health 2017 May 25;59(3):215-236 [FREE Full text] [doi: 10.1539/joh.16-0227-RA] [Medline: 28320977]
24. Weisel KK, Lehr D, Heber E, Zarski AC, Berking M, Riper H, et al. Severely burdened individuals do not need to be excluded from internet-based and mobile-based stress management: effect modifiers of treatment outcomes from three randomized controlled trials. J Med Internet Res 2018 Jun 19;20(6):e211 [FREE Full text] [doi: 10.2196/jmir.9387] [Medline: 29921562]
25. Schulz A, Stolz T, Vincent A, Krieger T, Andersson G, Berger T. A sorrow shared is a sorrow halved? A three-arm randomized controlled trial comparing internet-based clinician-guided individual versus group treatment for social anxiety disorder. Behav Res Ther 2016 Dec;84:14-26. [doi: 10.1016/j.brat.2016.07.001] [Medline: 27423374]
26. Freedman SA, Dayan E, Kimelman YB, Weissman H, Eitan R. Early intervention for preventing posttraumatic stress disorder: an internet-based virtual reality treatment. Eur J Psychotraumatol 2015;6:25608 [FREE Full text] [doi: 10.3402/ejpt.v6.25608] [Medline: 25843345]
27. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, de Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
28. Ashford RD, Lynch K, Curtis B. Technology and social media use among patients enrolled in outpatient addiction treatment programs: cross-sectional survey study. J Med Internet Res 2018 Mar 6;20(3):e84 [FREE Full text] [doi: 10.2196/jmir.9172] [Medline: 29510968]
29. Barnes S, Prescott J. Empirical evidence for the outcomes of therapeutic video games for adolescents with anxiety disorders: systematic review. JMIR Serious Games 2018 Feb 28;6(1):e3 [FREE Full text] [doi: 10.2196/games.9530] [Medline: 29490893]
30. Kouwenhoven-Pasmooij TA, Robroek SJ, Ling SW, van Rosmalen J, van Rossum EF, Burdorf A, et al. A blended web-based gaming intervention on changes in physical activity for overweight and obese employees: influence and usage in an

experimental pilot study. JMIR Serious Games 2017 Apr 3;5(2):e6 [FREE Full text] [doi: 10.2196/games.6421] [Medline: 28373157]

31. Boendermaker WJ, Veltkamp RC, Peeters M. Training behavioral control in adolescents using a serious game. Games Health J 2017 Dec;6(6):351-357. [doi: 10.1089/g4h.2017.0071] [Medline: 28926286]
32. Xu X, Li J, Pham TP, Salmon CT, Theng Y. Improving psychosocial well-being of older adults through exergaming: the moderation effects of intergenerational communication and age cohorts. Games Health J 2016 Dec;5(6):389-397. [doi: 10.1089/g4h.2016.0060] [Medline: 27976953]
33. Souders DJ, Boot WR, Charness N, Moxley JH. Older adult video game preferences in practice: investigating the effects of competing or cooperating. Games Cult 2016;11(1-2):170-120 [FREE Full text] [doi: 10.1177/1555412015603538] [Medline: 29033698]
34. Belchior P, Marsiske M, Sisco SM, Yam A, Bavelier D, Ball K, et al. Video game training to improve selective visual attention in older adults. Comput Human Behav 2013 Jul 1;29(4):1318-1324 [FREE Full text] [doi: 10.1016/j.chb.2013.01.034] [Medline: 24003265]
35. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
36. Morton TA, Wilson N, Haslam C, Birney M, Kingston R, McCloskey L. Activating and guiding the engagement of seniors with online social networking: experimental findings from the AGES 2.0 project. J Aging Health 2018 Jan;30(1):27-51. [doi: 10.1177/0898264316664440] [Medline: 27530332]
37. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
38. Rus HM, Tiemensma J. Social media as a shield: Facebook buffers acute stress. Physiol Behav 2018 Mar 1;185:46-54. [doi: 10.1016/j.physbeh.2017.12.021] [Medline: 29273455]
39. Mota PJ. Facebook enhances antidepressant pharmacotherapy effects. ScientificWorldJournal 2014;2014:892048 [FREE Full text] [doi: 10.1155/2014/892048] [Medline: 24574930]
40. Zhu Y, Zhang H, Tian M. Molecular and functional imaging of internet addiction. Biomed Res Int 2015;2015:378675 [FREE Full text] [doi: 10.1155/2015/378675] [Medline: 25879023]
41. Meerloo JA. Television addiction and reactive apathy. J Nerv Ment Dis 1954;120(3-4):290-291. [Medline: 13233976]
42. Sussman S, Moran MB. Hidden addiction: television. J Behav Addict 2013 Sep;2(3):125-132 [FREE Full text] [doi: 10.1556/jba.2.2013.008] [Medline: 25083294]
43. Griffiths MD, Kuss DJ, King DL. Video game addiction: past, present and future. Curr Psychiatry Rev 2012 Oct 1;8(4):308-318 [FREE Full text] [doi: 10.2174/157340012803520414]
44. Griffiths M. Does internet and computer. Cyberpsychol Behav 2000 Apr 21;3(2):211-218. [doi: 10.1089/109493100316067] [Medline: 5637250]
45. Young K, Pistner M, O'Mara J, Buchanan J. Cyber disorders: the mental health concern for the new millennium. Cyberpsychol Behav 1999;2(5):475-479. [doi: 10.1089/cpb.1999.2.475] [Medline: 19178220]
46. Young KS. Internet addiction: a new clinical phenomenon and its consequences. Am Behav Sci 2016 Jul 27;48(4):402-415 [FREE Full text] [doi: 10.1177/0002764204270278]
47. Kuss DJ, Kanjo E, Crook-Rumsey M, Kibowski F, Wang GY, Sumich A. Problematic mobile phone use and addiction across generations: the roles of psychopathological symptoms and smartphone use. J Technol Behav Sci 2018;3(3):141-149 [FREE Full text] [doi: 10.1007/s41347-017-0041-3] [Medline: 30238057]
48. Pera A. Psychopathological processes involved in social comparison, depression, and envy on Facebook. Front Psychol 2018 Dec;9(12):22 [FREE Full text] [doi: 10.3389/fpsyg.2018.00022] [Medline: 29410643]
49. Appel H, Gerlach AL, Crusius J. The interplay between Facebook use, social comparison, envy, and depression. Curr Opin Psychol 2016 Jun;9:44-49. [doi: 10.1016/j.copsyc.2015.10.006]
50. Pawlowski C. Glued To The Tube: The Threat Of Television Addiction To Today's Family. Naperville, Illinois: Sourcebooks Inc; 2000.
51. Fisher S. Identifying video game addiction in children and adolescents. Addict Behav 1994;19(5):545-553. [doi: 10.1016/0306-4603(94)90010-8] [Medline: 7832013]
52. Grüsser SM, Thalemann R, Griffiths MD. Excessive computer game playing: evidence for addiction and aggression? Cyberpsychol Behav 2007 Apr;10(2):290-292. [doi: 10.1089/cpb.2006.9956] [Medline: 17474848]
53. Nalwa K, Anand AP. Internet addiction in students: a cause of concern. Cyberpsychol Behav 2003 Dec;6(6):653-656. [doi: 10.1089/109493103322725441] [Medline: 14756932]
54. Stein DJ. Internet addiction, internet psychotherapy. Am J Psychiatry 1997 Jun;154(6):890. [doi: 10.1176/ajp.154.6.890] [Medline: 9167536]
55. Leung L. Predicting internet risks: a longitudinal panel study of gratifications-sought, internet addiction symptoms, and social media use among children and adolescents. Health Psychol Behav Med 2014 Jan 1;2(1):424-439 [FREE Full text] [doi: 10.1080/21642850.2014.902316] [Medline: 25750792]

56. Longstreet P, Brooks S. Life satisfaction: a key to managing internet & social media addiction. Technol Soc 2017 Aug;50:73-77. [doi: 10.1016/j.techsoc.2017.05.003]
57. Elhai JD, Hall BJ, Erwin MC. Emotion regulation's relationships with depression, anxiety and stress due to imagined smartphone and social media loss. Psychiatry Res 2018 Mar;261:28-34. [doi: 10.1016/j.psychres.2017.12.045] [Medline: 29276991]
58. Wartberg L, Kammerl R, Rosenkranz M, Hirschhäuser L, Hein S, Schwinge C, et al. The interdependence of family functioning and problematic internet use in a representative quota sample of adolescents. Cyberpsychol Behav Soc Netw 2014 Jan;17(1):14-18. [doi: 10.1089/cyber.2012.0494] [Medline: 23952627]
59. Yan W, Li Y, Sui N. The relationship between recent stressful life events, personality traits, perceived family functioning and internet addiction among college students. Stress Health 2014 Feb;30(1):3-11. [doi: 10.1002/smi.2490] [Medline: 23616371]
60. Yu L, Shek DT. Internet addiction in Hong Kong adolescents: a three-year longitudinal study. J Pediatr Adolesc Gynecol 2013 Jun;26(3 Suppl):S10-S17. [doi: 10.1016/j.jpag.2013.03.010] [Medline: 23683821]
61. Heo J, Oh J, Subramanian SV, Kim Y, Kawachi I. Addictive internet use among Korean adolescents: a national survey. PLoS One 2014;9(2):e87819 [FREE Full text] [doi: 10.1371/journal.pone.0087819] [Medline: 24505318]
62. Kalaitzaki AE, Birtchnell J. The impact of early parenting bonding on young adults' internet addiction, through the mediation effects of negative relating to others and sadness. Addict Behav 2014 Mar;39(3):733-736. [doi: 10.1016/j.addbeh.2013.12.002] [Medline: 24368006]
63. Tang J, Yu YZ, Du YK, Ma Y, Zhang DY, Wang JJ. Prevalence of internet addiction and its association with stressful life events and psychological symptoms among adolescent internet users. Addict Behav 2014 Mar;39(3):744-747. [doi: 10.1016/j.addbeh.2013.12.010] [Medline: 24388433]
64. Thomée S, Härenstam A, Hagberg M. Computer use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Psychiatry 2012 Oct 22;12:176 [FREE Full text] [doi: 10.1186/1471-244X-12-176] [Medline: 23088719]
65. Griffiths MD, Szabo A. Is excessive online usage a function of medium or activity? An empirical pilot study. J Behav Addict 2014 Mar;3(1):74-77 [FREE Full text] [doi: 10.1556/JBA.2.2013.016] [Medline: 25215216]
66. Pontes HM, Szabo A, Griffiths MD. The impact of internet-based specific activities on the perceptions of internet addiction, quality of life, and excessive usage: a cross-sectional study. Addict Behav Rep 2015 Jun;1:19-25 [FREE Full text] [doi: 10.1016/j.abrep.2015.03.002] [Medline: 29531976]
67. Pak R, Price MM, Thatcher J. Age-sensitive design of online health information: comparative usability study. J Med Internet Res 2009;11(4):e45 [FREE Full text] [doi: 10.2196/jmir.1220] [Medline: 19917549]
68. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
69. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U] [Medline: 7726811]
70. Baruffol E, Gisle L, Corten P. Life satisfaction as a mediator between distressing events and neurotic impairment in a general population. Acta Psychiatr Scand 1995 Jul;92(1):56-62. [doi: 10.1111/j.1600-0447.1995.tb09543.x] [Medline: 7572249]
71. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 1966;80(1):1-28. [doi: 10.1037/h0092976] [Medline: 5340840]
72. van der Eijk Y, Uusitalo S. Towards a 'Sociorelational' approach to conceptualizing and managing addiction. Public Health Ethics 2016 Jul;9(2):198-207 [FREE Full text] [doi: 10.1093/phe/phw013] [Medline: 27551305]
73. Henden E, Melberg HO, Røgeberg OJ. Addiction: choice or compulsion? Front Psychiatry 2013;4:77 [FREE Full text] [doi: 10.3389/fpsyt.2013.00077] [Medline: 23966955]
74. Hasebrink U, Popp J. Media repertoires as a result of selective media use: a conceptual approach to the analysis of patterns of exposure. Communications 2006;31(3):369-387. [doi: 10.1515/COMMUN.2006.023]
75. Landau SF. The relationship between objective and subjective social stress indicators: some Israeli findings. Eur Sociol Rev 1988 Dec;4(3):249-262. [doi: 10.1093/oxfordjournals.esr.a036488]
76. Donnelly K, Donnelly JP, Warner GC, Kittleson CJ, King PR. Longitudinal study of objective and subjective cognitive performance and psychological distress in OEF/OIF veterans with and without traumatic brain injury. Clin Neuropsychol 2018 Apr;32(3):436-455. [doi: 10.1080/13854046.2017.1390163] [Medline: 29052464]
77. Senn TE, Walsh JL, Carey MP. The mediating roles of perceived stress and health behaviors in the relation between objective, subjective, and neighborhood socioeconomic status and perceived health. Ann Behav Med 2014 Oct;48(2):215-224 [FREE Full text] [doi: 10.1007/s12160-014-9591-1] [Medline: 24648016]
78. Rönnlund M, Sundström A, Sörman DE, Nilsson LG. Effects of perceived long-term stress on subjective and objective aspects of memory and cognitive functioning in a middle-aged population-based sample. J Genet Psychol 2013;174(1):25-41. [doi: 10.1080/00221325.2011.635725] [Medline: 23534095]

79. Rehkopf DH, Kuper H, Marmot MG. Discrepancy between objective and subjective measures of job stress and sickness absence. Scand J Work Environ Health 2010 Nov;36(6):449-457 [FREE Full text] [Medline: 20725704]
80. Sussman S, Sussman AN. Considering the definition of addiction. Int J Environ Res Public Health 2011 Dec;8(10):4025-4038 [FREE Full text] [doi: 10.3390/ijerph8104025] [Medline: 22073026]
81. Young KS. Caught in the Net: How to Recognize the Signs of Internet Addiction--and a Winning Strategy for Recovery. New York, US: John Wiley & Sons Inc; 1998.
82. Mason JW. A review of psychoendocrine research on the pituitary-adrenal cortical system. Psychosom Med 1968;30(5):Suppl:576-Suppl:607 [FREE Full text] [doi: 10.1097/00006842-196809000-00020] [Medline: 4303377]
83. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
84. Schoorl J, Rijn SV, Wied MD, van Goozen S, Swaab H. The role of anxiety in cortisol stress response and cortisol recovery in boys with oppositional defiant disorder/conduct disorder. Psychoneuroendocrinology 2016 Dec;73:217-223. [doi: 10.1016/j.psyneuen.2016.08.007] [Medline: 27521740]
85. Vreeburg SA, Hoogendijk WJ, DeRijk RH, van Dyck R, Smit JH, Zitman FG, et al. Salivary cortisol levels and the 2-year course of depressive and anxiety disorders. Psychoneuroendocrinology 2013 Sep;38(9):1494-1502 [FREE Full text] [doi: 10.1016/j.psyneuen.2012.12.017] [Medline: 23313277]
86. Renna ME, O'Toole MS, Spaeth PE, Lekander M, Mennin DS. The association between anxiety, traumatic stress, and obsessive-compulsive disorders and chronic inflammation: a systematic review and meta-analysis. Depress Anxiety 2018 Nov;35(11):1081-1094. [doi: 10.1002/da.22790] [Medline: 30199144]
87. Thakur ER, Sansgiry S, Petersen NJ, Stanley M, Kunik ME, Naik AD, et al. Cognitive and perceptual factors, not disease severity, are linked with anxiety in COPD: results from a cross-sectional study. Int J Behav Med 2018 Dec;25(1):74-84. [doi: 10.1007/s12529-017-9663-2] [Medline: 28779469]
88. Biondi M, Picardi A. Psychological stress and neuroendocrine function in humans: the last two decades of research. Psychother Psychosom 1999;68(3):114-150. [doi: 10.1159/000012323] [Medline: 10224513]
89. Hughes C. Medicare annual wellness visits made easier. Fam Pract Manag 2011;18(4):10-14 [FREE Full text] [Medline: 21842820]
90. van Praag B, Frijters P, Ferrer-i-Carbonell A. The anatomy of subjective well-being. J Econ Behav Organ 2003 May;51(1):29-49. [doi: 10.1016/S0167-2681(02)00140-3]
91. Perrin A. Pew Research Centre October. 2015. Social Networking Usage: 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2019-01-22] [WebCite Cache ID 75clnYQde]
92. Gruzd A, Jacobson J, Mai P, Dubois E. RULA Digital Repository. 2018. The State of Social Media in Canada 2017 URL: https://digital.library.ryerson.ca/islandora/object/RULA%3A6432
93. Ryan T, Chester A, Reece J, Xenos S. The uses and abuses of Facebook: A review of Facebook addiction. J Behav Addict 2014 Sep;3(3):133-148 [FREE Full text] [doi: 10.1556/JBA.3.2014.016] [Medline: 25317337]
94. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
95. Escobar-Viera CG, Whitfield DL, Wessel CB, Shensa A, Sidani JE, Brown AL, et al. For better or for worse? A systematic review of the evidence on social media use and depression among lesbian, gay, and bisexual minorities. JMIR Ment Health 2018 Jul 23;5(3):e10496 [FREE Full text] [doi: 10.2196/10496] [Medline: 30037786]
96. Kuss DJ, Griffiths MD. Online social networking and addiction--a review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]
97. Fox J, Tokunaga RS. Romantic partner monitoring after breakups: attachment, dependence, distress, and post-dissolution online surveillance via social networking sites. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):491-498. [doi: 10.1089/cyber.2015.0123] [Medline: 26348808]
98. Sagioglou C, Greitemeyer T. Facebook's emotional consequences: why Facebook causes a decrease in mood and why people still use it. Comput Human Behav 2014 Jun;35:359-363. [doi: 10.1016/j.chb.2014.03.003]
99. Kim DJ, Kim K, Lee H, Hong J, Cho MJ, Fava M, et al. Internet game addiction, depression, and escape from negative emotions in adulthood: a nationwide community sample of Korea. J Nerv Ment Dis 2017 Jul;205(7):568-573. [doi: 10.1097/NMD.0000000000000698] [Medline: 28598958]
100. Loton D, Borkoles E, Lubman D, Polman R. Video game addiction, engagement and symptoms of stress, depression and anxiety: the mediating role of coping. Int J Ment Health Addiction 2015 Jul 31;14(4):565-578. [doi: 10.1007/s11469-015-9578-6]
101. Becker GS, Murphy KM. A theory of rational addiction. J Polit Econ 1988 Aug;96(4):675-700. [doi: 10.1086/261558]
102. van der Vijgh B, Beun R, van Rood M, Werkhoven P. Meta-analysis of digital game and study characteristics eliciting physiological stress responses. Psychophysiology 2015 Aug;52(8):1080-1098. [doi: 10.1111/psyp.12431] [Medline: 25950613]
103. Hébert S, Béland R, Dionne-Fournelle O, Crête M, Lupien SJ. Physiological stress response to video-game playing: the contribution of built-in music. Life Sci 2005 Apr 1;76(20):2371-2380. [doi: 10.1016/j.lfs.2004.11.011] [Medline: 15748630]

104. Griffiths M. Addiction on the internet or addiction to the internet?-the case of online gambling addiction. J Behav Addict 2013;2(Suppl 1):16. [doi: 10.1556/JBA.2.2013.Suppl.1]
105. Zarski AC, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Dec 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]

Abbreviations

ANOVA: analysis of variance
ES: emotional stress
IAT: Internet Addiction test
ICTs: information and communication technologies
SMS: short message service
VIF: variance inflation factor

Edited by G Eysenbach; submitted 09.07.18; peer-reviewed by L Becker, D Hasson; comments to author 08.10.18; revised version received 05.11.18; accepted 09.12.18; published 02.04.19.

Please cite as:
Khalili-Mahani N, Smyrnova A, Kakinami L
To Each Stress Its Own Screen: A Cross-Sectional Survey of the Patterns of Stress and Various Screen Uses in Relation to Self-Admitted Screen Addiction
J Med Internet Res 2019;21(4):e11485
URL: https://www.jmir.org/2019/4/e11485/
doi:10.2196/11485
PMID:30938685

Original Paper

The Relationship Between Personality Traits, Psychopathological Symptoms, and Problematic Internet Use: A Complex Mediation Model

Beatrix Koronczai[1*], PhD; Gyöngyi Kökönyei[2,3,4*], PhD; Mark D Griffiths[5*], PhD; Zsolt Demetrovics[2*], PhD, DSC

[1]Department of Developmental and Clinical Child Psychology, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary
[2]Department of Clinical Psychology and Addiction, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary
[3]SE-NAP2 Genetic Brain Imaging Migraine Research Group, Semmelweis University, Budapest, Hungary
[4]Department of Pharmacodynamics, Faculty of Pharmacy, Semmelweis University, Budapest, Hungary
[5]Psychology Division, Nottingham Trent University, Nottingham, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Beatrix Koronczai, PhD
Department of Developmental and Clinical Child Psychology
Institute of Psychology
Eötvös Loránd University
46 Izabella St
Budapest, H-1064
Hungary
Phone: 36 1 461 2600
Fax: 36 1 461 2697
Email: koronczai.beatrix@ppk.elte.hu

Abstract

Background: There are many empirical studies that demonstrate the associations between problematic internet use, psychopathological symptoms, and personality traits. However, complex models are scarce.

Objective: The aim of this study was to build and test a mediation model based on problematic internet use, psychopathological symptoms, and personality traits.

Methods: Data were collected from a medical addiction center (43 internet addicts) and internet cafés (222 customers) in Beijing (mean age 22.45, SD 4.96 years; 239/265, 90.2% males). Path analysis was applied to test the mediation models using structural equation modeling.

Results: Based on the preliminary analyses (correlations and linear regression), two different models were built. In the first model, low conscientiousness and depression had a direct significant influence on problematic internet use. The indirect effect of conscientiousness—via depression—was nonsignificant. Emotional stability only affected problematic internet use indirectly, via depressive symptoms. In the second model, low conscientiousness also had a direct influence on problematic internet use, whereas the indirect path via the Global Severity Index was again nonsignificant. Emotional stability impacted problematic internet use indirectly via the Global Severity Index, whereas it had no direct effect on it, as in the first model.

Conclusions: Personality traits (ie, conscientiousness as a protective factor and neuroticism as a risk factor) play a significant role in problematic internet use, both directly and indirectly (via distress level).

(J Med Internet Res 2019;21(4):e11837) doi:10.2196/11837

KEYWORDS

problematic internet use; personality; psychopathology; psychopathological symptoms

Introduction

Most empirical studies to date have found a positive association between problematic internet use and psychopathological symptoms in normal samples of both adolescents [1-5] and adults [6-11]. A few studies have examined this relationship among clinical samples (ie, among diagnosed internet addicts), comparing them to healthy control groups [2,12,13] or clinical control groups [14,15]. The results of sampling from both clinical and normal populations have demonstrated an increased level of psychopathological symptoms among problematic internet users. When predictor variables have been examined for problematic internet use, findings have also been consistent. In most studies, depressive [1,14,10,12,13,5,15] and obsessive-compulsive symptoms [8,1,9,14,13,4,15] have been found to be the most significant predictors of problematic internet use.

Additionally, several studies have reported important predictors of problematic internet use (or they are present at a more extensive level in the group of problematic internet users), including hostility [1,9,13,4,5], anxiety [10,12,5,13], and interpersonal sensitivity [8,1,15]. One longitudinal study [16] has provided indicative data concerning the cause-and-effect between problematic internet use and psychopathological symptoms. The results suggested that obsessive-compulsive symptoms are predictors of internet addiction, whereas increased levels of depression, anxiety, hostility, interpersonal sensitivity, and psychoticism are consequences of internet addiction.

A meta-analysis by Kayis et al [17] that evaluated 12 studies found that all five main factors of the Big Five model correlated with problematic internet use. More specifically, agreeableness, openness to experience, extraversion, and conscientiousness were negatively associated with internet addiction, whereas neuroticism was positively associated with internet addiction. In general, the relationship between neuroticism and problematic internet use appears the most established. Neuroticism has been positively associated with (1) problematic internet use in all empirical research to date in correlational studies (eg, [18-20]), (2) comparison of groups of internet addicts and controls (eg, [21,22]), and (3) regression analyses (eg, [23,19]). This association is also found in research assessing neuroticism by questionnaires based on (1) Eysenck's three-factor theory (eg, [24-33]) and (2) Zuckerman's five-factor model (eg, [34]). Similarly, studies have also reported an association between low agreeableness and internet addiction (eg, [21,23,18,20]) and low conscientiousness and internet addiction (eg, [18,22,20]).

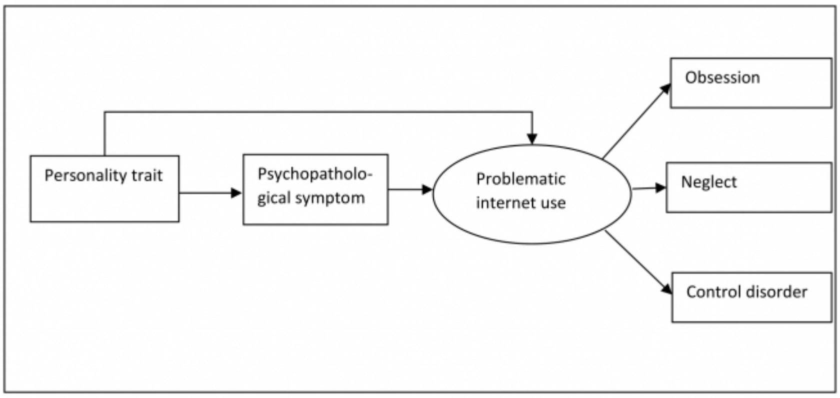
The direction of the association between extraversion and problematic internet use is controversial. Some studies have demonstrated a positive relationship with more symptoms of internet addiction associated with higher extraversion (eg, [21,18,20]). However, another study reported a negative association with a higher level of problematic internet use correlated with higher introversion [22]. Regarding Eysenck's three-factor model, introversion has also been related to problematic internet use in some cases (eg, [35,25,13,31]). Additionally, Zuckerman's sociability and activity factors (which may correspond with extraversion), have also been found to correlate negatively with internet addiction [34]. Similar incoherence has been found in the case of openness to experience. One study reported an association between problematic internet use and low openness to experience [21], whereas another reported a positive association between internet addiction and openness to experience [21].

To date, there have been relatively few mediation or moderation models examining the complex associations and interactions between personality traits, internet addiction, and other variables. Researchers have examined the associations between specific personality traits and problematic internet use via coping strategies [21]. Additionally, personality traits have been shown to mediate the impact of time spent online on internet addiction [36]. Kuss et al [23] also demonstrated that the interactions between different online activities and personality traits affect the likelihood of becoming an internet addict.

To the authors' knowledge, only two studies have tested complex models including variables comprising personality, psychopathology, and problematic internet use. One of them [37] presented a model in which personality was characterized in terms of the behavioral inhibition and behavioral activation systems, and depression, impulsivity, and anxiety were considered psychopathologies. The study found that both personality variables influenced internet addiction and that the effect was mediated by anxiety and/or depression and/or impulsivity in different ways. Floros et al [38] described a path model analysis in which personality traits were conceptualized by Zuckerman's alternative five-factor model, and psychopathological symptoms were assessed using the global indexes of the 90-item Symptom Checklist. In this model, personality traits and defense style both had an effect on internet addiction, and internet addiction predicted psychopathological symptoms (versus the reverse).

In summary, there are many empirical studies that demonstrate associations between internet addiction and psychopathological symptoms, and between internet addiction and personality traits. However, further analysis is needed on the complex effects and models. Given the lack of research, the aim of this study was to build and test a mediation model that examined personality factors, psychopathological symptoms, and problematic internet use within a single complex model (see Figure 1). The investigation of complex effects is relevant in particular for problematic internet users because the outcomes might facilitate the focus of their treatment.

Figure 1. Proposed mediation model.



Methods

Participants and Procedure

The data for this sample were collected from two samples of intensive internet users. Although the two samples appear to be distinct, this sampling method can be explained by specific Chinese circumstances. In an internet addiction clinic, the patients are not a simple treatment-seeking population because the young internet users often are delegated (and sometimes forced) to enter treatment by their parents. Based on some prior reports (eg, [39,40]), intensive users with a high risk for problematic internet use can be found in internet cafés. Sample 1 (the clinical group) consisted of diagnosed internet addicts who were hospitalized at an addiction medical center in Beijing that specializes in the treatment of problematic internet users. Each patient admitted to the hospital and diagnosed for problematic internet use was included in the sample during the 9 months of the study. In the case of patients younger than 18 years, both the patients and their parents were informed about the study goals and were asked to provide informed consent. Participation was voluntary, and the questionnaires were completed anonymously. Sample 2 (the internet café group) consisted of customers of internet cafés in the Chaoyang District of Beijing. Managers of 15 internet cafés were asked for permission to carry out the data collection, and 13 agreed. Each of the 13 cafés were visited three times. During data collection, each customer was invited to participate in the study and approximately 10% (222/1850) agreed to participate. A small gift was offered as recompense for participation in the study (ie, money for 2-hour internet use or a soft drink; appoximately US $1.50) was offered. The customers completed the questionnaires on site but via an online survey. Participation in the research was voluntary and anonymous. The participants could read information about the study and provide informed consent prior to completing the questionnaire. The study protocol was approved by the Institutional Review Board of Eötvös Loránd University, Budapest. The final sample consisted of 43 diagnosed internet addicts (42 males, 1 female) and 222 internet café customers (197 males, 25 females).

Measures

Demographic Data and Internet Use Characteristics

Basic personal demographic information and other questions were asked about the location, the duration, the frequency, and the purpose of the participants' internet use.

Problematic Internet Use Questionnaire-9

The Chinese version of Problematic Internet Use Questionnaire (PIUQ-9) [41] consists of three factors (obsession, neglect, and control disorder) with three items relating to each factor. The obsession subscale relates to mental withdrawal symptoms caused by the lack of internet use (eg, "How often do you feel tense, irritated, or stressed if you cannot use the internet for as long as you want to?"). The neglect subscale contains items related to difficulties in controlling internet use (eg, "How often do you spend time online when you'd rather sleep?"). The control disorder subscale relates to difficulties in controlling internet use (eg, "How often do you try to conceal the amount of time spent online?"). Participants use a 5-point Likert scale to estimate the extent to which each given statement is true to them. The scale ranges from 9 to 45; the maximum scores are 15 for the subscales. Higher scores indicate more symptoms of problematic internet use.

Big Five Mini-Markers

The Big Five Mini-Markers scale (BFI) [42] is a shortened version of Goldberg's unipolar Big-Five Markers [43] and consists of 40 adjectives. Participants evaluate every adjective according to how well it describes them on a 9-point Likert scale. It has five factors that assess the participants' overall personality (ie, extraversion, agreeableness, conscientiousness, emotional stability, and intellect/openness). In all subscales, higher scores indicate a higher level of that specific personality characteristic. The maximum score on each subscale is 72.

Brief Symptom Inventory

The Brief Symptom Inventory (BSI) test [44] is a shortened version of the Symptom Checklist-90-R [45]. It consists of 53 items; participants assess how much the symptoms bothered them the previous week on a 5-point Likert scale. The scale lists the clinically relevant psychological symptoms that are indicators of emotional distress. The items include nine dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoia, and psychoticism. For all the subscales, higher scores indicate more psychopathological symptoms. In addition, a global index was used, namely the Global Severity Index, which is the mean of all the items. The maximum score on the interpersonal sensitivity subscale is 20; 25 on the hostility, phobic anxiety, paranoid ideation, and psychoticism subscales; 30 on the obsessive-compulsive symptoms, depression, and anxiety subscales; 35 on the somatization subscale, and 5 on the Global Severity Index.

Statistical Analysis

For statistical analyses, SPSS version 23.0 and Mplus version 7.11 statistical software packages were used. In addition to the mean and standard deviation of the scales, Cronbach alphas were calculated as indexes of internal consistency, which were considered good if the values were at least .70 [46]. Correlational analysis and regression analysis were also applied. Based on these results, path analysis was used to test the mediation models with structural equation modeling using maximum likelihood estimation robust to nonnormality [47]. To evaluate the overall fit of the models, the absolute fit index (chi-square test), the comparative fit index (CFI), the Tucker-Lewis index (TLI) or nonnormed fit index, and the root mean square error approximation (RMSEA) were used. The CFI and TLI are related to the total variance accounted by the model, with values higher than 0.95 indicating a good fit, and values below 0.90 indicating a poor fit [48]. The RMSEA is related to the variance of the residuals, and values below 0.08 are considered an acceptable fit, while values below .05 indicate a good fit. Closeness of model fit (CFit) using RMSEA (CFit of RMSEA) evaluating the statistical deviation of RMSEA from the value 0.05 is also reported. Nonsignificant probability values (P>.05) indicate acceptable fit. However, some methodologists suggest values larger than P>.50 [48].

Results

Descriptive Statistics

The mean age of participants was 22.45 (SD 4.96) years in the total sample, 17.9 (SD 0.42) years in the clinical group, and 23.47 (SD 4.77) years in the internet café group. The age difference between the two samples was statistically significant (t_{217}=10.056, P<.001). The time spent on the internet for the purpose of studying or working is presented in Table 1. Approximately one-third of the sample used the internet for studying or working 3 to 4 hours a day. This represented the largest category out of the six options given among all internet use. Approximately 10% of the participants declared that they spent more than 8 hours a day online for the purpose of studying or working. Table 1 also shows the time spent on the internet for purposes other than studying or working and the pattern was similar. Two-thirds of the participants used the internet for entertainment 1 to 2 hours or 3 to 4 hours a day, and slightly less than 10% used the internet for entertainment for more than 8 hours a day.

The clinical group reported higher total PIUQ score and higher scores on the neglect factor than the internet café group. Also, a significant difference was found between the clinical group and the internet café group according to BFI intellect/openness (see Table 2). The effect size for differences in the total PIUQ score and for the neglect factor was small (Cohen *d*=0.41), but medium and large for the PIUQ neglect factor (Cohen *d*=0.64) and intellect/openness (Cohen *d*=0.87).

Correlations between the variables of the study are reported in Multimedia Appendix 1.

Based on previous results [49], 22 points (out of 45) was defined as a cut-off point of the PIUQ-9, which created two categories of internet users (problematic and nonproblematic users) The proportion of problematic internet users was 37% (16/43) in the clinical group and 31.9% (71/222) in the internet café group. Applying linear regression, symptoms which remained in significant relationships were tested with problematic internet use (which was a continuous variable) after controlling for the effects on one another. In addition to the sample category that the participants were in, the increased levels of obsessive-compulsive and depressive symptoms contributed significantly to an explanation of the variance of total scores (see Table 3).

Based on the preliminary analyses (correlations and linear regression), a model was built to investigate the relationships between problematic internet use, personality traits, and psychopathological symptoms (see Figure 2). It was assumed that depressive and obsessive-compulsive symptoms mediated the relationship between personality traits (emotional stability, conscientiousness) and problematic internet use (defined here as a latent variable). The subsample variable was also added to the model because there was a difference between the two subsamples in the PIUQ total score. Additionally, after performing linear regression, the subsample variable was significant in predicting the PIUQ score.

Table 1. Time spent on the internet for working/studying and other purposes (N=265).

Hours per day	Working or studying, n (%) (n=262)	Other purposes, n (%) (n=261)
<1	59 (22.5)	37 (14.2)
1-2	52 (19.8)	82 (31.4)
3-4	73 (27.9)	81 (31.0)
5-6	26 (9.9)	27 (10.3)
7-8	21 (8.0)	11 (4.2)
>8	31 (11.8)	23 (8.8)

Table 2. Means (standard deviations) and differences by group with Cronbach alphas.

Scale[a]	Cronbach alpha	Total sample (N=265), mean (SD)	Clinical group (n=43), mean (SD)	Internet café group (n=222), mean (SD)	*t* value (df)	*P* value
PIUQ-9						
Total	.848	20.10 (8.16)	23.15 (9.75)	19.53 (7.73)	2.223 (252)	.03
Obsession	.749	5.74 (3.09)	6.28 (3.55)	5.63 (2.98)	1.262 (260)	.21
Neglect	.713	7.20 (3.14)	8.95 (3.47)	6.87 (3.01)	3.966 (256)	<.001
Control	.886	7.12 (3.03)	7.79 (3.67)	7.00 (2.88)	1.554 (258)	.12
BSI						
Somatization	.840	9.80 (3.95)	9.71 (3.59)	9.81 (4.03)	0.157 (243)	.88
Obsessive-compulsive	.817	10.77 (4.54)	10.29 (4.07)	10.86 (4.64)	0.730 (241)	.47
Interpersonal sensitivity	.791	7.38 (3.59)	6.93 (3.09)	7.47 (3.68)	0.884 (243)	.38
Depression	.871	10.11 (5.00)	9.66 (3.86)	10.21 (5.20)	0.777 (243)	.44
Anxiety	.826	8.51 (3.75)	8.88 (3.84)	8.44 (3.74)	0.688 (243)	.49
Hostility	.790	7.85 (3.39)	8.22 (3.23)	7.78 (3.42)	0.759 (243)	.45
Phobic anxiety	.712	7.13 3.04)	6.88 (2.53)	7.18 (3.14)	0.582 (244)	.56
Paranoid ideation	.772	7.68 (3.30)	7.74 (3.26)	7.67 (3.32)	0.140 (245)	.89
Psychoticism	.775	7.73 (3.46)	8.07 (3.68)	7.66 (3.42)	0.694 (243)	.49
Global Severity Index	.970	1.57 (0.60)	1.56 (0.54)	1.58 (0.61)	0.145 (239)	.89
BFI						
Extraversion	.540	43.95 (7.87)	44.98 (10.83)	43.74 (7.14)	0.707 (247)	.48
Agreeableness	.711	51.24 (8.79)	53.02 (9.73)	50.88 (8.57)	1.441 (247)	.15
Conscientiousness	.712	44.20 (8.68)	44.69 (10.35)	44.10 (8.33)	0.349 (247)	.73
Emotional Stability	.734	45.61 (9.78)	47.05 (12.60)	45.32 (9.12)	0.845 (249)	.40
Intellect/Openness	.734	46.03 (8.87)	52.43 (9.33)	44.73 (8.20)	5.412 (247)	<.001

[a] BFI: Big Five Inventory; BSI: Brief Symptom Inventory; PIUQ-9: Problematic Internet Use Questionnaire.

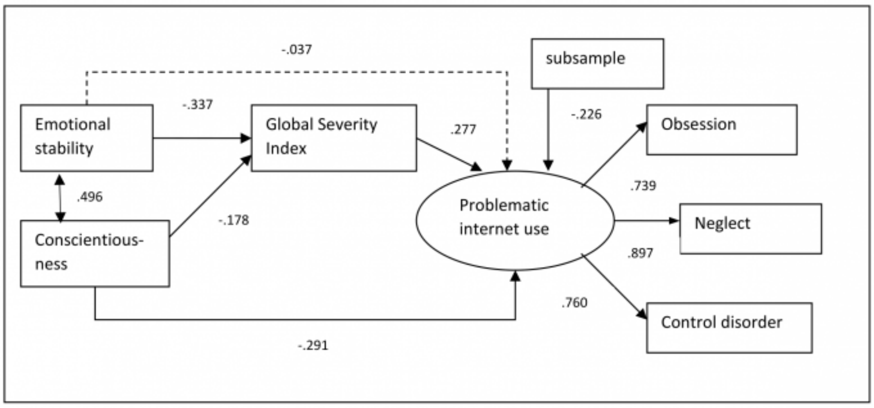
Table 3. Linear regression for prediction of problematic internet use (R^2=.239).

Independent variable	Standardized β	*P* value
Sex (0: male, 1: female)	–.013	.83
Subsample (0: clinical, 1: internet café)	–.198	.001
BSI[a]		
Somatization	–.023	.82
Obsessive-compulsive	.258	.01
Interpersonal sensitivity	–.020	.85
Depression	.362	.003
Anxiety	.005	.97
Hostility	.096	.34
Phobic anxiety	–.029	.79
Paranoid ideation	–.130	.22
Psychoticism	–.087	.47

[a] BSI: Brief Symptom Inventory.

Figure 2. The mediation model and standardized path coefficients. Dashed arrows indicate nonsignificant path coefficients; continuous arrows indicate significant paths.

Figure 3. Mediation model with Global Severity Index. Dashed arrows indicate nonsignificant path coefficients; continuous arrows indicate significant paths.



The goodness-of-fit indexes of the mediation model were appropriate (χ^2_{14}=14.5, *P*=.28; CFI=0.995, TLI=0.991, RMSEA=0.026, 90% CI 0.000-0.068, CFit=0.792). Low conscientiousness and depression had a direct significant effect on problematic internet use, whereas the direct effects of emotional stability and obsessive-compulsive symptoms were nonsignificant. Both emotional stability and low conscientiousness significantly explained the symptoms of depression and obsessive-compulsive disorder. This meant that low conscientiousness directly impacted problematic internet use. However, the indirect effect of low conscientiousness—via depression—was nonsignificant (standardized indirect effect=–0.047, *P*=.11). Emotional stability only affected problematic internet use indirectly, via depressive symptoms (standardized indirect effect=–0.059, *P*=.03). The impact of the sample category on problematic internet use was significant (see Figure 2). The participants in the clinical sample had higher scores on the PIUQ-9 compared to the internet café sample. The model explained 32.5% of the total variance of problematic internet use. Given that all the psychopathological symptoms positively correlated with problematic internet use, another mediation model was tested in which the Global Severity Index was used instead of the individual symptoms (see Figure 3).

The goodness-of-fit indexes of the second mediation model were good (χ^2_{11}=16.2, *P*=.13; CFI=0.985, TLI=0.975, RMSEA=0.042, 90% CI: 0.000-0.083). Low conscientiousness had a direct effect on problematic internet use, whereas the indirect path via the Global Severity Index was nonsignificant (standardized indirect effect=–0.049, *P*=.10). Emotional stability impacted problematic internet use indirectly via the Global Severity Index (standardized indirect effect=–0.094, *P*<.001), whereas it had no direct effect on it. The model explained 28.9% of the total variance of problematic internet use.

Discussion

The results of this study demonstrated that both samples showed much higher levels of problematic internet use than those observed in normal populations (7.1% in Asia) [50]. Although this was expected in the clinical sample, the similar prevalence among those recruited from internet cafés was nonevident at first sight. However, internet cafés have a special position in Chinese internet culture [51-54]. In internet cafés, young people (mostly males, younger than 30 years) play online games, chat online, and watch movies. It is perhaps not surprising that the prevalence of internet addiction is high among the patronage of internet cafés [39,40,55]. Furthermore, Griffiths et al [56] noted that parents in Southeast Asian countries appear to pathologize any behavior of their children that takes time away from educational pursuits and family. This tendency—the parents tend to feel anxious due to their (mainly male) children’s school performance—might lead to more vigilance for any symptoms of problematic internet use and to seeking help for their adolescents. Psychiatrists interviewed the problematic internet users in this study; however, the diagnosis of internet addiction is not official, and the scale used to assess problematic internet use is not based on official diagnostic criteria. Consequently, there might be a discrepancy in the level of symptoms based on currently used clinical interviews and the scale used in this study (PIUQ).

Based on the outcomes of the preliminary statistical analyses, low conscientiousness and emotional stability negatively correlated with problematic internet use. These findings are congruent with previous results reported in the literature on problematic internet use [21,23,18,19,22,20]. In their meta-analysis, Kotov et al [57] found in adults that high neuroticism (equivalent to low emotional stability) and low conscientiousness were also associated with anxiety, depression, and substance use disorders.

Neuroticism was the strongest correlate among the five personality traits, and low conscientiousness was the second trait to have a strong and consistently negative effect size. In another study [58], similar findings were reported. Extraversion, low conscientiousness, and low emotional stability had the strongest predictive values on psychopathological symptoms. In a large sample of psychosomatic outpatients [59], the level of neuroticism was a differentiating factor between the clinical and nonclinical samples with a large effect size. Additionally, patients with higher neuroticism and low conscientiousness were more likely to have a personality disorder. Therefore, it appears that the importance of these two personality traits is not specific to problematic internet use but is common in psychopathologies more generally.

The other three personality traits of the Big Five (ie, agreeableness, openness, and extraversion) did not correlate with problematic internet use in the sample in this study. This result might be explained by the fact that the recruited sample was very specific, including a higher proportion of users with more severe problems. Thus, it is tempting to hypothesize that emotional instability and low conscientiousness might be those personality factors that contribute to the maintenance of problematic internet use. However, prospective studies are needed to test this notion. In addition, it is worth noting that the previous correlational findings between problematic internet, openness, and extraversion were mixed, thus further studies are needed utilizing different samples.

Among the psychopathological symptoms, only obsessive-compulsive symptoms and depression were significant predictors of problematic internet use. These findings are in line with previous results [8,1,14,12,13,4,5,15]. In reviewing other addictive behaviors, there are some additional findings that reinforce the results of this study. For instance, in the case of compulsive buying, Maraz et al [60] found an increased level of obsessive-compulsive symptoms among addicted shoppers compared to nonaddicted shoppers. Moussas et al [61] investigated patients of a methadone maintenance treatment program, and depression and obsessive-compulsive symptoms were found to have the highest mean scores among all the symptoms. Similarly, in the case of methamphetamine users, obsessive-compulsive symptoms and depression were reported to have the highest levels among the psychopathological symptoms, especially for injectors (compared to methamphetamine users who used other routes of administration) [62]. Based on the previously mentioned findings, the association between problematic internet use and specific psychopathological symptoms is similar to the associations between other addictive behaviors and specific psychopathological symptoms (obsessive-compulsive symptoms and/or depression).

The correlational analyses showed that all the psychopathological symptoms correlated with problematic internet use (r=.268-.404). Additionally, using the Global Severity Index, the mediation model corresponded with the data. In this second model, the path coefficient of Global Severity Index to problematic internet use was higher compared to that of the individual symptoms in the first model. Overall, it appears that the level of psychological distress (as indicated by the Global Severity Index) is a more important factor regarding problematic internet use than the specificity of psychopathology.

Based on fit indexes, both models showed excellent fit to the data. Because the two models were not nested, they could not be compared directly. However, results of the two models appear to be convergent. More specifically, emotional stability only affected problematic internet use indirectly via psychopathological symptoms (regardless of the indexes used), whereas low conscientiousness only had a direct effect on problematic internet use.

The first mediation model examined in this study was partly in line with previous findings. According to Smits and Boeck [63], the behavioral inhibition system relates to neuroticism. In Park et al's [37] mediation model, the behavioral inhibition system impacted internet addiction via depression, which reinforces the findings of the model here, low emotional stability had an indirect effect on problematic internet use (however, the direct effect was also significant). Regarding low conscientiousness, which negatively relates to the BASF (ie, the fun-seeking scale of the Behavioral Activation System) [63], Park et al's study also found a direct association between BASF and internet addiction, similar to the findings of this study (between low conscientiousness and problematic internet use). However, in their model, the indirect effect was significant in the case of impulsiveness and anxiety, whereas this study did not show any significant indirect effects between conscientiousness, depression, and obsessive-compulsive symptoms. Based on the outcome of the second path analysis, it could be concluded that low emotional stability only affects problematic internet use indirectly (via psychological distress), whereas low conscientiousness affects problematic internet use directly.

Interpreting the models proposed here, two different types of problematic users might be considered in terms of personality. Problematic internet use has long been known as a heterogeneous phenomenon [64]. Chamberlain et al [65] found that problematic internet use exists with and without other impulsive/compulsive conditions. However, both impair quality of life. It might be assumed that there are different paths leading to problematic internet use depending on the user's personality. One path could be when an individual with a high level of neuroticism tries to cope with their negative emotions by repeatedly using the internet more intensively (ie, compensatory internet use [66]). In such cases, the level of psychological distress (eg, depressive feelings) mediates between neuroticism and problematic internet use. Because neuroticism is associated (prospectively) to internalizing symptoms [67], a possible path from neuroticism into problematic internet could be via internalizing symptoms (ie, depression and anxiety).

The other path could be when an individual with a low level of conscientiousness becomes vulnerable to problematic internet use. Low conscientiousness is regarded as being disorganized, inefficient, careless, and sloppy because these characteristics equate to a deficit in the executive functions. This could also provide an explanation for the comorbidity with attention-deficit/hyperactivity disorder (ADHD) [68-70]. This theory is reinforced by previously reported findings. For example, Van Dijk et al [71] found that adults with ADHD showed a higher level of neuroticism and a lower level of conscientiousness than healthy controls. Additionally, Gomez and Corr [72] reported in their meta-analysis that inattentional symptoms were associated with low conscientiousness. Regarding internet gaming disorder (IGD), Argyriou et al [73] also conducted a meta-analysis and demonstrated that there was an association between IGD and impaired response inhibition. They conceptualized IGD as externalizing psychopathology. This is in line with Dong and Potenza's [74] suggestion of a cognitive-behavioral model of IGD.

It should also be noted that obsessive-compulsive symptoms were assessed by items such as trouble remembering things, difficulty making decisions, and trouble concentrating. These items might also signal a deficit in the executive functions. However, this subscale was not a significant mediator variable between low conscientiousness and problematic internet use. In future research, it would be worth investigating impulsivity rather than obsessive-compulsive symptoms in the model, such as Park et al’s [37], or assessing executive functions with cognitive tests (eg, inhibitory control, decision making, shifting).

Nevertheless, in the model proposed here, the two paths were not independent from each other. This fact is consistent with other results and theories on different executive functions and the internalizing-externalizing dichotomy. Executive functions may also be divided into hot and cool components [75], in which hot executive functions are involved in highly motivating and emotional situations. Based on this differentiation, neuroticism is associated with the executive function [76,77]. Additionally, there is evidence that component facets of neuroticism and conscientiousness share a common neurological system, in which high neuroticism and low conscientiousness associate with lower scores on the executive function battery [78]. Similarly, internalizing and externalizing disorders are not independent from each other either [79]. Additionally, depression is associated not only with neuroticism but also with conscientiousness [80]. Hall et al [81] noted the role of both personality (primary conscientiousness and neuroticism) and executive functions in predicting health behavior patterns, which might underpin the relevance of the model presented here. However, the models only explained 32.5% and 28.9% of the variances of the PIUQ. Consequently, further research is needed to identify other important factors shaping the symptoms of problematic internet use. In addition to users’ individual personalities, situational, social, and environmental factors would also be worth investigating.

One of the major implications of the findings in this study is that clinicians should be educated about the possible cultural aspects regarding the associations of personality traits, psychopathological symptoms, and problematic internet use. Additionally, the findings of this study highlight the possibility of the differences between internet users concerning intensity of usage in the role of personal characteristics in developing problematic internet use.

Finally, it should be noted that this study has several limitations. First, the sample was nonrepresentative of internet users and included intensive internet users. More representative samples are needed in any replication. The sample was Chinese only and may not be representative of internet users in other countries. Therefore, future research should also include participants of other countries and cultures. The sample size was modest (although adequate for the statistical testing carried out) and future studies should try to recruit as large a sample as possible. It is also suggested that future studies should include samples with a more even distribution of females because the sample in this study was predominantly male. Gender differences can then be explored more thoroughly. Finally, the data were self-reported and open to well-known biases (eg, social desirability and poor memory recall). Taking these limitations together, generalization of the findings should be applied with caution. To gain reliable data, more objective reports should be added (eg, family members’ and friends’ reports on the internet user’s behaviors).

In conclusion, our study revealed the role personality plays in problematic internet use. However, to clarify the associations between different personality traits and internet addiction, further investigations are necessary that apply complex models including possible mediator variables such as psychopathological symptoms.

Acknowledgments

This study was supported by Science and Technology Fellowship Programme China, funded by the European Union and the Hungarian National Research, Development and Innovation Office (grant numbers: K111938, KKP126835). Gyöngyi Kökönyei and Beatrix Koronczai acknowledge the financial support of the János Bolyai Research Fellowship awarded by the Hungarian Academy of Sciences. Beatrix Koronczai is also supported by the ÚNKP-18-4 New National Excellence Program of the Ministry of Human Capacities. The funding institutions had no role in the study design or collection, analysis, and interpretation of the data; in writing the manuscript; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Correlations between the PIUQ and the subscales of BSI and BFI.

[PDF File (Adobe PDF File), 83KB - jmir_v21i4e11837_app1.pdf]

References

1. Jang KS, Hwang SY, Choi JY. Internet addiction and psychiatric symptoms among Korean adolescents. J Sch Health 2008 Mar;78(3):165-171. [doi: 10.1111/j.1746-1561.2007.00279.x] [Medline: 18307612]
2. Müller KW, Dreier M, Duven E, Giralt S, Beutel ME, Wölfling K. Adding clinical validity to the statistical power of large-scale epidemiological surveys on internet addiction in adolescence: a combined approach to investigate psychopathology

and development-specific personality traits associated with internet addiction. J Clin Psychiatry 2017 Mar;78(3):e244-e251. [doi: 10.4088/JCP.15m10447] [Medline: 28199069]
3. Siomos K, Floros G, Makris E, Christou G, Hadjulis M. Internet addiction and psychopathology in a community before and during an economic crisis. Epidemiol Psychiatr Sci 2014 Sep;23(3):301-310. [doi: 10.1017/S2045796013000437] [Medline: 23962651]
4. Yang C, Choe B, Baity M, Lee J, Cho J. SCL-90-R and 16PF profiles of senior high school students with excessive internet use. Can J Psychiatry 2005 Jun;50(7):407-414. [doi: 10.1177/070674370505000704] [Medline: 16086538]
5. Yen J, Ko C, Yen C, Chen S, Chung W, Chen C. Psychiatric symptoms in adolescents with Internet addiction: comparison with substance use. Psychiatry Clin Neurosci 2008 Feb;62(1):9-16 [FREE Full text] [doi: 10.1111/j.1440-1819.2007.01770.x] [Medline: 18289136]
6. Alavi SS, Alaghemandan H, Maracy MR, Jannatifard F, Eslami M, Ferdosi M. Impact of addiction to internet on a number of psychiatric symptoms in students of isfahan universities, iran, 2010. Int J Prev Med 2012 Feb;3(2):122-127 [FREE Full text] [Medline: 22347609]
7. Beranuy M, Oberst U, Carbonell X, Chamarro A. Problematic Internet and mobile phone use and clinical symptoms in college students: the role of emotional intelligence. Comp Hum Behav 2009 Sep;25(5):1182-1187. [doi: 10.1016/j.chb.2009.03.001]
8. Dalbudak E, Evren C, Topcu M, Aldemir S, Coskun KS, Bozkurt M, et al. Relationship of Internet addiction with impulsivity and severity of psychopathology among Turkish university students. Psychiatry Res 2013 Dec 30;210(3):1086-1091. [doi: 10.1016/j.psychres.2013.08.014] [Medline: 23998359]
9. Laconi S, Kaliszewska-Czeremska K, Gnisci A, Sergi I, Barke A, Jeromin F, et al. Cross-cultural study of problematic Internet use in nine European countries. Comp Hum Behav 2018 Jul;84:430-440. [doi: 10.1016/j.chb.2018.03.020]
10. Ghaderi Rammazi M, Askarizadeh G, Ahmadi G, Divsalar K. The role of psychiatric symptoms, social support and meaning in life in predicting internet addiction among university students: a causal model. PCP 2018 Mar 01;6(2):101-110. [doi: 10.29252/nirp.jpcp.6.2.101]
11. Taymur I, Budak E, Demirci H, Akdağ HA, Güngör BB, Özdel K. A study of the relationship between internet addiction, psychopathology and dysfunctional beliefs. Comp Hum Behav 2016 Aug;61:532-536. [doi: 10.1016/j.chb.2016.03.043]
12. Tonioni F, D'Alessandris L, Lai C, Martinelli D, Corvino S, Vasale M, et al. Internet addiction: hours spent online, behaviors and psychological symptoms. Gen Hosp Psychiatry 2012;34(1):80-87. [doi: 10.1016/j.genhosppsych.2011.09.013] [Medline: 22036735]
13. Xiuqin H, Huimin Z, Mengchen L, Jinan W, Ying Z, Ran T. Mental health, personality, and parental rearing styles of adolescents with Internet addiction disorder. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):401-406. [doi: 10.1089/cyber.2009.0222] [Medline: 20712498]
14. Müller KW, Beutel ME, Wölfling K. A contribution to the clinical characterization of Internet addiction in a sample of treatment seekers: validity of assessment, severity of psychopathology and type of co-morbidity. Compr Psychiatry 2014 May;55(4):770-777. [doi: 10.1016/j.comppsych.2014.01.010] [Medline: 24602498]
15. Wölfling K, Beutel ME, Koch A, Dickenhorst U, Müller KW. Comorbid internet addiction in male clients of inpatient addiction rehabilitation centers: psychiatric symptoms and mental comorbidity. J Nerv Ment Dis 2013 Nov;201(11):934-940. [doi: 10.1097/NMD.0000000000000035] [Medline: 24177479]
16. Dong G, Lu Q, Zhou H, Zhao X. Precursor or sequela: pathological disorders in people with Internet addiction disorder. PLoS One 2011 Feb 16;6(2):e14703 [FREE Full text] [doi: 10.1371/journal.pone.0014703] [Medline: 21358822]
17. Kayiş AR, Satici SA, Yilmaz MF, Şimşek D, Ceyhan E, Bakioğlu F. Big five-personality trait and internet addiction: a meta-analytic review. Comp Hum Behav 2016 Oct;63:35-40 [FREE Full text] [doi: 10.1016/j.chb.2016.05.012]
18. Kuss DJ, Shorter GW, van Rooij AJ, van de Mheen D, Griffiths MD. The Internet addiction components model and personality: establishing construct validity via a nomological network. Comp Hum Behav 2014 Oct;39:312-321. [doi: 10.1016/j.chb.2014.07.031]
19. Malik A, Rafiq N. Exploring the relationship of personality, loneliness, and online social support with internet addiction and procrastination. Pak J Psychol Res 2016;31(1):93-117 [FREE Full text]
20. Zhou Y, Li D, Li X, Wang Y, Zhao L. Big five personality and adolescent Internet addiction: the mediating role of coping style. Addict Behav 2017 Jan;64:42-48. [doi: 10.1016/j.addbeh.2016.08.009] [Medline: 27543833]
21. Kakulte A, Shejwal B. Internet addiction and personality differences among adolescents. Indian J Health Wellbeing 2017 Apr;8(4):315-318 [FREE Full text]
22. Müller KW, Koch A, Dickenhorst U, Beutel ME, Duven E, Wölfling K. Addressing the question of disorder-specific risk factors of internet addiction: a comparison of personality traits in patients with addictive behaviors and comorbid internet addiction. Biomed Res Int 2013;2013:546342 [FREE Full text] [doi: 10.1155/2013/546342] [Medline: 23865056]
23. Kuss DJ, Griffiths MD, Binder JF. Internet addiction in students: prevalence and risk factors. Comp Hum Behav 2013 May;29(3):959-966. [doi: 10.1016/j.chb.2012.12.024]
24. Cao F, Su L. Internet addiction among Chinese adolescents: prevalence and psychological features. Child Care Health Dev 2007 May;33(3):275-281. [doi: 10.1111/j.1365-2214.2006.00715.x] [Medline: 17439441]

25. Dalbudak E, Evren C. The relationship of Internet addiction severity with attention deficit hyperactivity disorder symptoms in Turkish university students: impact of personality traits, depression and anxiety. Compr Psychiatry 2014 Apr;55(3):497-503 [FREE Full text] [doi: 10.1016/j.comppsych.2013.11.018] [Medline: 24374171]
26. Dong G, Wang J, Yang X, Zhou H. Risk personality traits of Internet addiction: a longitudinal study of Internet-addicted Chinese university students. Asia Pac Psychiatry 2013 Dec;5(4):316-321 [FREE Full text] [doi: 10.1111/j.1758-5872.2012.00185.x] [Medline: 23857796]
27. Mehroof M, Griffiths M. Online gaming addiction: the role of sensation seeking, self-control, neuroticism, aggression, state anxiety, and trait anxiety. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):313-316 [FREE Full text] [doi: 10.1089/cyber.2009.0229] [Medline: 20557251]
28. Öztürk E, Özmen SK. The relationship of self-perception, personality and high school type with the level of problematic internet use in adolescents. Comp Hum Behav 2016 Dec;65:501-507. [doi: 10.1016/j.chb.2016.09.016]
29. Senormancı O, Saraçlı O, Atasoy N, Senormancı G, Koktürk F, Atik L. Relationship of Internet addiction with cognitive style, personality, and depression in university students. Compr Psychiatry 2014 Aug;55(6):1385-1390. [doi: 10.1016/j.comppsych.2014.04.025] [Medline: 24889340]
30. Tsai H, Cheng S, Yeh T, Shih C, Chen K, Yang Y, et al. The risk factors of Internet addiction--a survey of university freshmen. Psychiatry Res 2009 May 30;167(3):294-299. [doi: 10.1016/j.psychres.2008.01.015] [Medline: 19395052]
31. Yan W, Li Y, Sui N. The relationship between recent stressful life events, personality traits, perceived family functioning and internet addiction among college students. Stress Health 2014 Feb;30(1):3-11. [doi: 10.1002/smi.2490] [Medline: 23616371]
32. Yao M, He J, Ko D, Pang K. The influence of personality, parental behaviors, and self-esteem on Internet addiction: a study of Chinese college students. Cyberpsychol Behav Soc Netw 2014 Feb;17(2):104-110 [FREE Full text] [doi: 10.1089/cyber.2012.0710] [Medline: 24003966]
33. Wu C, Lee M, Liao S, Chang L. Risk factors of internet addiction among internet users: an online questionnaire survey. PLoS One 2015;10(10):e0137506 [FREE Full text] [doi: 10.1371/journal.pone.0137506] [Medline: 26462196]
34. Kumar P, Singh U. Internet addiction in relation to personality factors of Zuckerman's alternative five factor model. Indian J Health Wellbeing 2014 Apr;5(4):500-502 [FREE Full text]
35. Ebeling-Witte S, Frank M, Lester D. Shyness, internet use, and personality. Cyberpsychol Behav 2007 Oct;10(5):713-716. [doi: 10.1089/cpb.2007.9964] [Medline: 17927542]
36. Durak M, Senol-Durak E. Which personality traits are associated with cognitions related to problematic Internet use? Asian J Soc Psychol 2014 Aug 08;17(3):206-218. [doi: 10.1111/ajsp.12056]
37. Park SM, Park YA, Lee HW, Jung HY, Lee J, Choi J. The effects of behavioral inhibition/approach system as predictors of Internet addiction in adolescents. Pers Individ Dif 2013 Jan;54(1):7-11. [doi: 10.1016/j.paid.2012.07.033]
38. Floros G, Siomos K, Stogiannidou A, Giouzepas I, Garyfallos G. The relationship between personality, defense styles, internet addiction disorder, and psychopathology in college students. Cyberpsychol Behav Soc Netw 2014 Oct;17(10):672-676. [doi: 10.1089/cyber.2014.0182] [Medline: 25225916]
39. Lei H, Li S, Chiu MM, Lu M. Social support and Internet addiction among mainland Chinese teenagers and young adults: a meta-analysis. Comp Hum Behav 2018 Aug;85:200-209. [doi: 10.1016/j.chb.2018.03.041]
40. Jiang Q. Internet addiction among young people in China: Internet connectedness, online gaming, and academic performance decrement. Internet Res 2014 Jan 28;24(1):2-20. [doi: 10.1108/IntR-01-2013-0004]
41. Koronczai B, Kökönyei G, Urbán R, Király O, Nagygyörgy K, Felvinczi K, et al. Confirmation of the Chinese version of the Problematic Internet Use Questionnaire Short Form (PIUQ-SF). Int J Ment Health Addiction 2016 May 16;15(1):191-197. [doi: 10.1007/s11469-016-9664-4]
42. Saucier G. Mini-markers: a brief version of Goldberg's unipolar big-five markers. J Pers Assess 1994 Dec;63(3):506-516. [doi: 10.1207/s15327752jpa6303_8] [Medline: 7844738]
43. Goldberg LR. The development of markers for the Big-Five factor structure. Psychol Assess 1992;4(1):26-42. [doi: 10.1037/1040-3590.4.1.26]
44. Derogatis L. The Brief Symptom Inventory (BSI): Administration, Scoring and Procedures Manual. Minneapolis, MN: National Computer Systems; 1993.
45. Derogatis L. SCL-90-R: Administration, Scoring and Procedures: Manual II. Baltimore, MD: Clinical Psychometric Research; 1983.
46. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997 Feb 22;314(7080):572 [FREE Full text] [Medline: 9055718]
47. Muthén L, Muthén B. Mplus User's Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén; 1998.
48. Brown T. Confirmatory Factor Analysis For Applied Research. First Edition. New York: The Guilford Press; 2006.
49. Koronczai B, Urbán R, Kökönyei G, Paksi B, Papp K, Kun B, et al. Confirmation of the three-factor model of problematic internet use on off-line adolescent and adult samples. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):657-664 [FREE Full text] [doi: 10.1089/cyber.2010.0345] [Medline: 21711129]

50. Cheng C, Li Y. Internet addiction prevalence and quality of (real) life: a meta-analysis of 31 nations across seven world regions. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):755-760 [FREE Full text] [doi: 10.1089/cyber.2014.0317] [Medline: 25489876]
51. Hong J, Huang L. A split and swaying approach to building information society: the case of Internet cafes in China. Telemat Inform 2005 Nov;22(4):377-393. [doi: 10.1016/j.tele.2004.11.005]
52. Liu F. It is not merely about life on the screen: urban Chinese youth and the Internet café. J Youth Studies 2009 Mar 26;12(2):167-184. [doi: 10.1080/13676260802590386]
53. Qiu JL. Cybercafés in China: community access beyond gaming and tight government control. Libr Trends 2013;62(1):121-139. [doi: 10.1353/lib.2013.0021]
54. Szablewicz M. The ill effects of “opium for the spirit”: a critical cultural analysis of China's Internet addiction moral panic. Chin J Commun 2010 Dec;3(4):453-470. [doi: 10.1080/17544750.2010.516579]
55. Müller KW, Glaesmer H, Brähler E, Woelfling K, Beutel ME. Prevalence of internet addiction in the general population: results from a German population-based survey. Behav Inform Technol 2013 Aug 15;33(7):757-766. [doi: 10.1080/0144929X.2013.810778]
56. Griffiths MD, Kuss DJ, Billieux J, Pontes HM. The evolution of Internet addiction: a global perspective. Addict Behav 2016 Feb;53:193-195. [doi: 10.1016/j.addbeh.2015.11.001] [Medline: 26562678]
57. Kotov R, Gamez W, Schmidt F, Watson D. Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol Bull 2010 Sep;136(5):768-821. [doi: 10.1037/a0020327] [Medline: 20804236]
58. Mirnics Z, Heincz O, Bagdy G, Surányi Z, Gonda X, Benko A, et al. The relationship between the big five personality dimensions and acute psychopathology: mediating and moderating effects of coping strategies. Psychiatr Danub 2013 Dec;25(4):379-388 [FREE Full text] [Medline: 24247050]
59. Schmitz N, Hartkamp N, Baldini C, Rollnik J, Tress W. Psychometric properties of the German version of the NEO-FFI in psychosomatic outpatients. Pers Individ Dif 2001 Oct;31(5):713-722. [doi: 10.1016/S0191-8869(00)00173-2]
60. Maraz A, van den Brink W, Demetrovics Z. Prevalence and construct validity of compulsive buying disorder in shopping mall visitors. Psychiatry Res 2015 Aug 30;228(3):918-924. [doi: 10.1016/j.psychres.2015.04.012] [Medline: 26027442]
61. Moussas G, Fanouraki I, Pachi A, Asomatou A, Drylli O, Paschalakis G, et al. Comorbid psychopathology and alcohol use patterns among methadone maintenance treatment patients. J Addict 2015;2015:197652 [FREE Full text] [doi: 10.1155/2015/197652] [Medline: 25878919]
62. Zweben JE, Cohen JB, Christian D, Galloway GP, Salinardi M, Parent D, Methamphetamine Treatment Project. Psychiatric symptoms in methamphetamine users. Am J Addict 2004;13(2):181-190. [doi: 10.1080/10550490490436055] [Medline: 15204668]
63. Smits DJ, Boeck PD. From BIS/BAS to the big five. Eur J Pers 2006 Jun;20(4):255-270. [doi: 10.1002/per.583]
64. Griffiths M. Internet addiction-time to be taken seriously? Addict Res 2009 Jul 11;8(5):413-418 [FREE Full text] [doi: 10.3109/16066350009005587]
65. Chamberlain SR, Ioannidis K, Grant JE. The impact of comorbid impulsive/compulsive disorders in problematic Internet use. J Behav Addict 2018 Jun 01;7(2):269-275. [doi: 10.1556/2006.7.2018.30] [Medline: 29788748]
66. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. Comp Hum Behav 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
67. Mezquita L, Ibáñez MI, Villa H, Fañanás L, Moya-Higueras J, Ortet G. Five-factor model and internalizing and externalizing syndromes: a 5-year prospective study. Pers Individ Dif 2015 Jun;79:98-103. [doi: 10.1016/j.paid.2015.02.002]
68. Carli V, Durkee T, Wasserman D, Hadlaczky G, Despalins R, Kramarz E, et al. The association between pathological internet use and comorbid psychopathology: a systematic review. Psychopathology 2013;46(1):1-13. [doi: 10.1159/000337971] [Medline: 22854219]
69. Han DH, Kim SM, Bae S, Renshaw PF, Anderson JS. Brain connectivity and psychiatric comorbidity in adolescents with Internet gaming disorder. Addict Biol 2017 May;22(3):802-812. [doi: 10.1111/adb.12347] [Medline: 26689148]
70. Leménager T, Hoffmann S, Dieter J, Reinhard I, Mann K, Kiefer F. The links between healthy, problematic, and addicted Internet use regarding comorbidities and self-concept-related characteristics. J Behav Addict 2018 Mar 01;7(1):31-43 [FREE Full text] [doi: 10.1556/2006.7.2018.13] [Medline: 29444606]
71. Van Dijk FE, Mostert J, Glennon J, Onnink M, Dammers J, Vasquez AA, et al. Five factor model personality traits relate to adult attention-deficit/hyperactivity disorder but not to their distinct neurocognitive profiles. Psychiatry Res 2017 Dec;258:255-261. [doi: 10.1016/j.psychres.2017.08.037] [Medline: 28844557]
72. Gomez R, Corr PJ. ADHD and personality: a meta-analytic review. Clin Psychol Rev 2014 Jul;34(5):376-388. [doi: 10.1016/j.cpr.2014.05.002] [Medline: 24929793]
73. Argyriou E, Davison CB, Lee TTC. Response inhibition and Internet gaming disorder: a meta-analysis. Addict Behav 2017 Aug;71:54-60. [doi: 10.1016/j.addbeh.2017.02.026] [Medline: 28264786]
74. Dong G, Potenza MN. A cognitive-behavioral model of Internet gaming disorder: theoretical underpinnings and clinical implications. J Psychiatr Res 2014 Nov;58:7-11 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.07.005] [Medline: 25062755]

75. Zelazo P, Müller U. Executive function in typical and atypical development. In: Goswami U, editor. Blackwell Handbook of Childhood Cognitive Development. Oxford: Blackwell Handbooks of Developmental Psychology; 2002:445-469.
76. Robinson MD. Personality, affective processing, and self-regulation: toward process-based views of extraversion, neuroticism, and agreeableness. Social Pers Psych Compass 2007 Nov;1(1):223-235. [doi: 10.1111/j.1751-9004.2007.00019.x]
77. Robison MK, Gath KI, Unsworth N. The neurotic wandering mind: an individual differences investigation of neuroticism, mind-wandering, and executive control. Q J Exp Psychol (Hove) 2017 Apr;70(4):649-663. [doi: 10.1080/17470218.2016.1145706] [Medline: 26821933]
78. Forbes CE, Poore JC, Krueger F, Barbey AK, Solomon J, Grafman J. The role of executive function and the dorsolateral prefrontal cortex in the expression of neuroticism and conscientiousness. Soc Neurosci 2014;9(2):139-151. [doi: 10.1080/17470919.2013.871333] [Medline: 24405294]
79. Cosgrove VE, Rhee SH, Gelhorn HL, Boeldt D, Corley RC, Ehringer MA, et al. Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. J Abnorm Child Psychol 2011 Jan;39(1):109-123 [FREE Full text] [doi: 10.1007/s10802-010-9444-8] [Medline: 20683651]
80. Klein DN, Kotov R, Bufferd SJ. Personality and depression: explanatory models and review of the evidence. Annu Rev Clin Psychol 2011;7:269-295 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032210-104540] [Medline: 21166535]
81. Hall PA, Fong GT, Epp LJ. Cognitive and personality factors in the prediction of health behaviors: an examination of total, direct and indirect effects. J Behav Med 2014 Dec;37(6):1057-1068. [doi: 10.1007/s10865-013-9535-4] [Medline: 24072429]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
BSI: Brief Symptom Inventory
CFI: comparative fit index
IGD: internet gaming disorder
PIUQ: Problematic Internet Use Questionnaire
RMSEA: root mean square error approximation
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 06.08.18; peer-reviewed by W Su, M El Tantawi; comments to author 27.09.18; revised version received 19.12.18; accepted 03.02.19; published 26.04.19.

Please cite as:
Koronczai B, Kökönyei G, Griffiths MD, Demetrovics Z
The Relationship Between Personality Traits, Psychopathological Symptoms, and Problematic Internet Use: A Complex Mediation Model
J Med Internet Res 2019;21(4):e11837
URL: http://www.jmir.org/2019/4/e11837/
doi:10.2196/11837
PMID:31025955

Original Paper

Digital Technology for Internet Access by Patients With Early-Stage Schizophrenia in Spain: Multicenter Research Study

Patricia Fernández-Sotos[1,2], MD; Antonio Fernández-Caballero[2,3,4], PhD; Pascual González[2,3,4], PhD; Ana Isabel Aparicio[2,5], BSN; Isabel Martínez-Gras[1], MD, PhD; Iosune Torio[1,2,6], MD; Mónica Dompablo[1,2], PhD; Lorena García-Fernández[2,7], MD, PhD; José Luis Santos[2,5], MD, PhD; Roberto Rodriguez-Jimenez[1,2,8], MD, PhD

[1]Department of Psychiatry, Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Madrid, Spain
[2]CIBERSAM (Biomedical Research Networking Centre in Mental Health), Madrid, Spain
[3]Instituto de Investigación en Informática, Albacete, Spain
[4]Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete, Spain
[5]Servicio de Psiquiatría del “Hospital Virgen de la Luz”, Cuenca, Spain
[6]Universidad Rey Juan Carlos, Madrid, Spain
[7]Hospital Universitario San Juan, Alicante, Spain
[8]CogPsy-Group, Universidad Complutense de Madrid, Madrid, Spain

Corresponding Author:
Roberto Rodriguez-Jimenez, MD, PhD
Department of Psychiatry
Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12)
Avda. de Córdoba, s/n
Madrid, 28041
Spain
Phone: 34 913908426
Email: roberto.rodriguez.jimenez@gmail.com

Abstract

Background: Digital technology and social networks are part of everyday life in the current internet age, especially among young people. To date, few studies have been published worldwide on the pattern of use of digital technology devices and applications in patients with early-stage schizophrenia and even fewer comparing them with healthy participants (not using data from general population surveys) from the same demographic areas. In Spain, no such study has been carried out.

Objective: The aim of this study was to analyze how patients with early-stage schizophrenia use internet and social networks compared with healthy participants matched by age and gender and also to examine which devices are utilized to access internet resources.

Methods: A cross-sectional, multicentric study was carried out through a semistructured interview asking about the use of digital technology devices and internet. The sample comprised 90 patients and 90 healthy participants. The semistructured interview was conducted on 30 outpatients and 30 healthy subjects in each of the 3 different cities (Madrid, Alicante, and Cuenca). Student *t* test was used for continuous variables and chi-square test for categorical variables. In the case of ordinal variables, nonparametric Mann-Whitney *U* and Kruskal-Wallis *H* tests for independent samples were performed to compare groups.

Results: The results indicated that a large proportion of patients with early-stage schizophrenia have access to different digital devices and use them frequently. In addition, both groups coincide in the order of preference and the purpose for which they use the devices. However, a lower frequency of use of most digital technology devices was detected in patients compared with healthy participants. In the case of some devices, this was due to the impossibility of access and not a lack of interest.

Conclusions: To our knowledge, this is the first study to analyze patterns of internet access and use of digital technology devices and applications in Spanish patients with early-stage schizophrenia compared with healthy participants from the same demographic areas. The results on significant access and use of digital technology and internet shown in this cross-sectional study will allow enhanced and more efficient treatment strategies to be planned, utilizing digital technology devices, for patients with early-stage schizophrenia.

(J Med Internet Res 2019;21(4):e11824) doi:10.2196/11824

KEYWORDS

information technology; computers; internet; schizophrenia

Introduction

Digital Technology

As indicated in a very recent report by the International Telecommunication Union (ITU) and the United Nations Educational, Scientific and Cultural Organization, the growth in information and communication technologies and the massive use of internet in the last 20 years have influenced the development of human activity in multiple areas such as education and health [1]. In this sense, internet access has increased worldwide over the last 10 years, reaching 48% of the population in 2017; in the case of Europe, the ITU reports a statistic of 79.6%. Facebook seems to be the favorite application throughout the world, surpassing 2 billion active users per month and reaching 1320 million active users per day in June 2017 (of whom, approximately 91% access Facebook via mobile technologies). WhatsApp and YouTube follow closely with around 1 billion users in June 2017 [1].

Moreover, in 2017, according to the Instituto Nacional de Estadística (INE; Spanish National Institute of Statistics), 78.4% of households with at least one member, aged between 16 and 74 years, had a computer and 97.4% of households had a landline or mobile phone [2]. In addition, 8 out of 10 participants between the ages of 16 and 74 years had used the internet in the previous 3 months and 2 out of every 3 had done so daily, with the use being even more frequent among young people. In the previous 3 months, 98% of young people aged between 16 and 24 years had used the internet, 91.3% on a daily basis and 49% for Web-based purchase. The products and services most commonly acquired were holiday accommodation (54.1%), sports equipment, clothing (53.5%), tickets for shows (47.6%), and other services for trips (44.7%). When analyzing the types of activities carried out on the internet by the Spanish population in 2017, those that were most frequently performed were receiving or sending emails; reading news, newspapers, or Web-based news magazines; searching for information about services; looking for information about health issues; and participating in social networks [2].

According to the same statistics by the INE, the device most commonly used to connect to the internet was by far the smartphone (90.4% of internet users in the last 3 months), followed by laptop (39.3%), desktop computer, and tablet. Smart television (TV), other mobile devices, and game consoles were also mentioned (12%). Regarding participation in social networks, during the previous 3 months, 67.6% of internet users had participated in general nature social networks such as Facebook or Twitter, creating a user profile or sending messages and other contributions, with greater participation by women than by men. The most participative subjects were students (90.4%) and people aged between 16 and 24 years (90.0%).

The fact that technology is rapidly changing society, and many activities now require the ability to use digital technology, potentially poses new problems for several population groups, including older adults, the economically disadvantaged, and people with severe mental illness [3]. Moreover, despite the clear potential of digital technology to connect people and health data in new ways, a key challenge is to ensure that patients and their needs remain at the center of technology development and implementation [4,5].

Use of Digital Technology in Health Care

In the last 10 years, the development and use of mobile devices devoted to health has increased significantly. The main advantage of these devices is that they increase access to medical care, reduce costs, and offer new options for control, prevention, detection of diseases, and basic diagnosis [6].

The literature at the time of this study reports positive effects derived from the use of mobile apps for health in relation to improving hygienic dietary habits such as stopping smoking, losing weight, dieting, and physical activity, as well as increasing therapeutic adherence and preventing and treating sexually transmitted diseases [7-9].

Use of Digital Technology in Mental Health

The impact of psychological interventions on mental disorders is unquestionable. However, owing to the limited resources of medical and psychological care and availability of interventions, their enormous potential is restricted [10-11]. Using internet and mobile-based interventions for mental disorders offer an accessible, innovative, and personalized option that addresses several of the devastating effects of mental illness, including associated stigma and the chronic nature and symptoms of these disorders. Patients are thus empowered to participate actively in their recovery [12]. Although digital exclusion among people with mental illness is still present, especially in older individuals with severe mental disorders, patients' access to internet-enabled technology is growing [13]. Internet and mobile-based studies or interventions have been shown to be effective in understanding and managing different mental disorders such as substance use disorder [14], depression [15], anxiety [16,17], and schizophrenia [18].

Use of Digital Technology in Schizophrenia and Related Disorders

Schizophrenia is a chronic psychiatric disorder that affects approximately 1% of the population and severely limits the social and occupational functioning of patients [19]. Symptoms of schizophrenia include positive symptoms (hallucinations, delusions, and disorganization of language and behavior), negative symptoms (including abulia, associability, anhedonia, and alogia or affective flattening), alterations in mood, and deficits in cognition [20].

Schizophrenia is a mental disorder for which mobile health (mHealth) offers a tremendous opportunity to provide personalized, innovative, and accessible solutions. A recent systematic review concluded that internet and mobile-based interventions for psychosis seem to be cost-effective, accessible, acceptable, feasible, and have the potential to improve clinical and social outcomes [21]. A subsequent review including studies carried out in 12 different countries supported the feasibility and acceptability of emerging mHealth and electronic health (eHealth) interventions among people with serious mental illness (including schizophrenia, schizoaffective disorder, and bipolar disorder) [22]. Recently, the need for individuals with schizophrenia to engage themselves in Web-based activities was demonstrated [23].

At the time of this study, there was limited information on the access and use of digital technology by psychotic patients. Several previous studies have investigated the prevalence of use of technological devices and applications through surveys. In the United States, 2 studies were conducted in patients diagnosed with schizophrenia and schizoaffective disorder. The first interviewed 457 participants by means of a Web-based survey [12] and the second surveyed 80 patients in Georgia [24]. In addition, 2 other studies were carried out to compare access and the use of internet in persons with different mental disorders, the first in France [25] and the second in Australia [26]. In Europe, 2 other studies examined access and purpose of internet use in patients with schizophrenia. The first study compared patients from Greece and Finland [27] and the second, patients from 2 psychiatric units in Finland [28].

To our knowledge, only 2 studies have been conducted in patients with early-stage schizophrenia. The first study included 71 patients at different stages of their 5-year treatment in a therapeutic program in Montreal (Canada). The patients completed a survey on their access and use of technology and related activities. The authors concluded that a significant proportion of patients with early-stage schizophrenia had access to and were using different technological devices in their daily lives [29]. The second study, carried out in Valencia (Spain), compared access, use, and interest in new technologies and eHealth interventions through a survey on internet use conducted on 65 patients with early psychosis compared with 40 patients with chronic psychosis [30].

It should be noted that in the previously mentioned publications, the use of digital technology for internet access was compared with data of use from general population surveys, but not with a sample of controls matched in age, gender, and place of residence. The lack of a control group, as well as a nonmulticentric study design, could be an important source of bias in these publications.

Objectives

To date, no multicentric studies have been carried out that provide information about the use of digital technology to access the internet by patients with early-stage schizophrenia compared with a sample of control participants matched in age, gender, and place of residence. Hence, we conducted a cross-sectional, multicentric study to analyze the use of digital technology in patients with early-stage schizophrenia and to compare the responses with regard to healthy participants matched by age, gender, and place of residence through a semistructured interview. The following 3 main questions were analyzed:

1. Which digital technology devices are most frequently used by patients with early-stage schizophrenia to access the internet? Is there any difference in use by gender, age, education level, or location? And, primarily, is use similar in healthy participants?
2. What is the main purpose of the use of digital technology devices in early-stage schizophrenia patients? Moreover, is there any difference in the purpose of use between patients and healthy participants?
3. Which are the applications more frequently used when patients access the internet? Is their use similar in healthy participants?

Methods

Study Design

This was a cross-sectional, multicentric study of 6 months' duration (June to November 2017). The design included 3 recruitment centers: *Hospital Universitario 12 de Octubre* (Madrid), *Hospital Virgen de la Luz* (Cuenca), and *Hospital Universitario de San Juan* (Alicante). The study was approved by the Clinical Research Ethics Committee.

The psychiatric service of the *Hospital Universitario 12 de Octubre* serves a population of about 450,000 inhabitants and has an integrated psychosis assistance program, including an intensive program for first psychotic episodes. The psychiatric service of the *Hospital Virgen de la Luz* serves a population of 150,000 inhabitants and also offers an integrated psychosis program. Finally, the psychiatric service of the *Hospital de San Juan* serves a population of 300,000 inhabitants and manages a program for first psychotic episodes.

Participants

The patient sample size was established as 90 patients (30 patients from each of the centers). All met the Diagnostic and Statistical Manual of Mental Disorders, fifth edition, (DSM-5) diagnostic criteria for schizophrenia assessed with the Structured Clinical Interview for DSM-5 [31]. All the patients were at an early stage of the disorder (5 years or less since their first episode). The mean time of evolution of the disorder was 2.6 years (SD 1.3). A total of 90 healthy participants were recruited (30 from each center), matched with patients for age and gender.

Table 1. Sociodemographic data of the samples.

Participants	Patients (n=90)	Healthy participants (n=90)	Statistics		
			Student *t* test (df)	Chi-square test	*P* value
Age (years), mean (SD)	28.1 (SD 8.6)	27.9 (SD 8.7)	0.189 (179)	—[a]	0.85
Gender (%)					
Men	53	57	—	0.2	0.65
Women	47	43	—	0.2	0.65
Educational level (%)					
Basic	29	22	—	8.4	0.02
Medium	50	37	—	8.4	0.02
High	21	41	—	8.4	0.02

[a] Not applicable.

Inclusion and Exclusion Criteria

The following inclusion criteria were established for patients:

1. Meeting DSM-5 diagnostic criteria for schizophrenia.
2. Staying clinically stabilized during the 3 months before the semistructured interview, according to criteria already used by our group [32].
3. Being an outpatient.
4. Being aged between 18 and 55 years.
5. Speaking fluent Spanish.
6. Providing signed, informed consent.

The following exclusion criteria were considered:

1. Other Axis I major mental disorders of DSM-5.
2. Intellectual disability (IQ <70).
3. Suffering somatic pathology that might interfere with accessing the internet.

The patients were screened by their psychiatrist to determine whether they met the inclusion/exclusion criteria and were thus suitable to participate in the study.

Table 1 shows the sociodemographic data of patients and controls. As expected, there were no differences in age or gender; the only differences were in educational level.

Data Collection Procedure

The patients were recruited consecutively at the clinical appointments of their respective first-episode programs. After the clinical evaluation, the psychiatrist assessed the inclusion and exclusion criteria and proposed their participation in the study on the frequency and purpose of using a series of digital technology devices. Multimedia Appendix 1 includes all the information on the data collection procedure considering the 32-item checklist of the Consolidated Criteria for Reporting Qualitative Research [33]. The healthy participants were recruited from areas of similar sociocultural status as the patients, mainly from similar cultural and social groups.

Statistical Analysis

Statistical analysis of all data was conducted using SPSS version 23.0 (IBM Corp). The means and SDs were used to describe continuous variables whereas percentages and chi-square tests were used for categorical variables. Student *t* test was used for continuous variables with a normal distribution. For variables found not to be normally distributed after using Levene test (of homogeneity of variance or homoscedasticity) and ordinal variables, nonparametric tests for independent samples were performed.

The tables described in the section on frequency of use include variables that are considered ordinal variables as they are clearly ordered from *never* (0), including both *never, but would like to* and *never, and would not like to*; *rarely* (1); *once a week* (2); *twice a week* (3); and *every day* (4). Therefore, the *P* value of nonparametric tests for independent samples is presented as a result. The Mann-Whitney *U* test (also called Mann-Whitney-Wilcoxon test or Wilcoxon rank sum test) is performed when we have 2 groups of variables whereas the Kruskal-Wallis *H* (Kruskal-Wallis one-way analysis of variance) test is conducted for more than 2 (in our case, always 3) groups of variables.

Results

The results for the technological devices used by patients and how they used them are presented. We analyze whether there were any differences for several features such as gender, age, education level, and place of residence. In addition, the use of these digital technology devices by patients with early-stage schizophrenia and healthy participants is compared. Subsequently, the purpose of use of these devices is shown, analyzing 4 different domains: entertainment, work, socialization, and shopping. In addition, the websites used to access these internet services and the main applications used for socialization are studied. Finally, the favorite videogames played by both groups are analyzed.

Frequency of Use of Digital Technology Devices

A first analysis studied the frequency of use by patients with early-stage schizophrenia of several technological devices that are useful for internet access. In this selection, some devices frequently used for this task such as computers or smartphones and other more novel or not so widely used devices (tablet, game console, or smart TV) were included. The inclusion of these latter devices allows possible trends in use to be evaluated.

Moreover, this variety of devices was included to cover different domains and usages.

Table 2 shows the frequency of use of these devices by patients with early-stage schizophrenia. In this table, not only the real frequency of use was included, taking values *every day*, *twice a week*, *once a week*, and *rarely*, but also in the case of *never* for a specific device, an analysis of patients' interest in their use (*never, but would like* and *never, and would not like*) was included. This variable reveals the actual motivation of low use of a specific device.

Considering the results shown in Table 2, the use of *smartphone* clearly prevails over other devices. More than 80% of patients (75 out of 90 participants) use this device *every day* and it is very popular among patients. The other type of device most frequently used is the computer. In this case, use is not so frequent but the general use of these devices by patients is similar to that of smartphones.

In contrast to the devices previously analyzed, the other 3 are used less frequently, but as can be seen in Table 2, patients' interest in their use is high.

Comparison Between Patients and Controls

First, we analyzed whether the frequency of use of these devices is similar between schizophrenic patients and healthy participants. As can be seen in Table 3, the frequency of use of most digital technology devices under study is higher in healthy participants than patients.

As can be seen, there is a significant statistical difference in almost all devices. Only the use of *game console* has a *P* value higher than .05.

Table 4 shows that there are no statistically significant differences between female and male patients in the use of digital technology devices. However, in general, females used *smartphone* and *computer* more frequently and, more remarkably, *game console* is used by a proportional number of each gender. In contrast, males were more frequent users of *smart TV* and *tablet* use was similar between genders.

Table 2. Frequency of use of digital technology devices in patients with early-stage schizophrenia (N=90).

Device	Every day, n (%)	Twice a week, n (%)	Once a week, n (%)	Rarely, n (%)	Never, but would like to, n (%)	Never, and would not like to, n (%)
Computer	50 (56)	17 (19)	9 (10)	7 (8)	5 (6)	2 (2)
Tablet	8 (9)	2 (2)	8 (9)	17 (19)	33 (37)	22 (24)
Smartphone	75 (83)	9 (10)	0 (0)	0 (0)	4 (4)	2 (2)
Game console	9 (10)	11 (12)	19 (21)	15 (17)	21 (23)	15 (17)
Smart TV	18 (20)	4 (4)	6 (7)	10 (11)	31 (34)	21 (23)

Table 3. Comparison between patients with early-stage schizophrenia (N=90) and healthy participants (N=90) using the Mann-Whitney U test on the frequency of use of digital technology devices.

Devices	Patients, mean (SD)	Healthy participants, mean (SD)	*P* value[a]
Computer	3.04 (1.306)	3.72 (0.750)	<.001
Tablet	0.79 (1.250)	1.22 (1.322)	.006
Smartphone	3.63 (1.022)	3.96 (0.422)	<.001
Game console	1.36 (1.376)	1.63 (1.532)	.23
Smart TV	1.18 (1.619)	1.97 (1.757)	.002

[a] Significance *P*<.05.

Table 4. Use of digital technology devices in patients with early-stage schizophrenia divided by gender using the Mann-Whitney *U* test (N=90).

Digital device	Gender		*P* value[a]
	Female (47%), mean (SD)	Male (53%), mean (SD)	
Computer	3.17 (1.305)	2.94 (1.311)	.30
Tablet	0.76 (1.306)	0.81 (1.283)	.97
Smartphone	3.81 (0.671)	3.48 (1.238)	.22
Game console	1.36 (1.376)	1.35 (1.391)	.93
Smart TV	0.88 (1.468)	1.44 (1.712)	.10

[a] Significance *P*<.05.

In terms of age (see Table 5), this study only found a slight statistically significant difference in the use of *smart TV*. Participants aged 30 years or above used *smart TV* less than younger adults. Although not statistically significant, a trend of greater use of most devices for the age range between 25 and 29 years can be observed across all participants. Nevertheless, it is worth highlighting that the age range in the last interval (≥30) is greater (30 to 53 years), which could affect the results.

For educational level (see Table 6), the only significant difference in terms of statistics is found for *computer*; the higher the educational level, the more frequent is the use. In general, participants with a high educational level used the analyzed devices more.

Table 7 shows some key details about the influence of the location where patients live. In this case, 2 different locations were analyzed, namely *rural* and *urban*. In general, the use of digital technology devices is greater in *urban* than in *rural* areas. There is, however, a statistically significant difference in the use of *game console*.

Table 5. Use of digital technology devices in patients with early-stage schizophrenia divided by age using the Kruskal-Wallis *H* test (N=90).

Digital device	Age (years)			*P* value[a]
	18-24 (37%), mean (SD)	25-29 (38%), mean (SD)	≥30 (25%), mean (SD)	
Computer	2.88 (1.386)	3.35 (0.884)	2.83 (1.642)	0.51
Tablet	0.67 (1.109)	0.97 (1.403)	0.70 (1.222)	0.53
Smartphone	3.58 (1.001)	3.82 (0.716)	3.43 (1.376)	0.26
Game console	1.18 (1.185)	1.74 (1.543)	1.04 (1.296)	0.16
Smart TV	1.30 (1.590)	1.53 (1.780)	0.48 (1.201)	0.03

[a] Significance *P*<.05.

Table 6. Use of digital technology devices in patients with early-stage schizophrenia divided by educational level using the Kruskal-Wallis *H* test (N=90).

Digital device	Educational level			*P* value[a]
	Basic (29%), mean (SD)	Medium (50%), mean (SD)	High (21%), mean (SD)	
Computer	2.50 (1.503)	3.04 (1.278)	3.79 (0.535)	.004
Tablet	0.77 (1.275)	0.62 (0.936)	1.21 (1.751)	.74
Smartphone	3.38 (1.299)	3.64 (1.026)	3.95 (0.230)	.14
Game console	1.23 (1.423)	1.24 (1.282)	1.79 (1.376)	.35
Smart TV	0.73 (1.218)	1.42 (1.764)	1.21 (1.686)	.37

[a] Significance *P*<.05.

Table 7. Use of digital technology devices in patients with early-stage schizophrenia divided by rural versus urban place of residence using the Mann-Whitney *U* test (N=90).

Digital device	Place of residence		*P* value[a]
	Rural (21%), mean (SD)	Urban (79%), mean (SD)	
Computer	3.11 (1.370)	3.03 (1.298)	.77
Tablet	0.47 (0.964)	0.87 (1.309)	.29
Smartphone	3.37 (1.499)	3.70 (0.852)	.93
Game console	0.89 (1.487)	1.48 (1.329)	.048
Smart TV	0.74 (1.522)	1.30 (1.634)	.08

[a] Significance *P*<.05.

Table 8. Comparison between patients with early-stage schizophrenia (N=90) and healthy participants (N=90) using the Mann-Whitney U test on the purpose of use of digital technology devices.

Participants	Patients, mean (SD)	Healthy participants, mean (SD)	*P* value[a]
Entertainment			
Computer	0.76 (0.430)	0.71 (0.457)	>.99
Tablet	0.74 (0.443)	0.80 (0.404)	.07
Smartphone	0.82 (0.385)	0.91 (0.288)	.02
Game console	0.74 (0.442)	0.95 (0.218)	.02
Smart TV	0.82 (0.393)	0.92 (0.281)	.003
Work			
Computer	0.63 (0.487)	0.81 (0.395)	.001
Tablet	0.23 (0.426)	0.16 (0.373)	.19
Smartphone	0.71 (0.454)	0.79 (0.412)	.10
Game console	0.04 (0.191)	0.07 (0.250)	.41
Smart TV	0.18 (0.393)	0.15 (0.363)	.60
Socialization			
Computer	0.58 (0.566)	0.61 (0.491)	.27
Tablet	0.23 (0.426)	0.20 (0.404)	.66
Smartphone	0.88 (0.326)	0.98 (0.149)	.003
Game console	0.52 (0.947)	0.25 (0.434)	.27
Smart TV	0.18 (0.393)	0.08 (0.281)	.11
Shopping			
Computer	0.47 (0.502)	0.69 (0.467)	.001
Tablet	0.11 (0.323)	0.22 (0.417)	.34
Smartphone	0.42 (0.496)	0.55 (0.500)	.05
Game console	0.02 (0.136)	0.07 (0.250)	.17
Smart TV	0.05 (0.226)	0.07 (0.254)	.71

[a] Significance $P<.05$.

Purpose of Use of Digital Technology Devices

The purpose of use was classified into 4 domains: entertainment, work, socialization, and shopping.

We analyzed the responses of participants who reported using a specific device. Again, an ordinal variable was assigned to the responses with the following Boolean values: NO=0 and YES=1. This means that the answers of participants who reported never using a specific device were not considered.

Table 8 shows the results based on the number of patients and healthy individuals who used the devices: *computer* was used by 83 patients versus 89 healthy individuals, *tablet* by 35 patients versus 55 healthy individuals, *smartphone* by 84 patients versus 89 healthy individuals, *game console* by 54 patients versus 61 healthy individuals, and *smart TV* by 38 patients versus 59 healthy individuals.

As can be seen, *shopping* is the least used domain for all devices. The data show that patients do not frequently shop on the internet. Another relevant result is that the more versatile devices are *computer* and *smartphone*, whereas the others are mainly used for *entertainment*. Regarding *socialization*, the high value associated with *game console* is curious. This outcome may show that the games played by patients involve relevant social activity. As expected, the use of *smartphone* is also noteworthy in this domain.

However, are there any differences between patients and healthy participants in the purpose of use? Table 8 also shows the results of a case-control study using the Mann-Whitney *U* test for nonparametric variables.

Regarding *entertainment*, the mean of healthy participants is higher in general terms and there are significant differences in terms of statistics in the case of *smartphone*, *game console*, and *smart TV*, where healthy participants used these devices for *entertainment* more than patients.

In the case of *work*, *computer* was also used more by healthy participants. It is noticeable, although not from a statistical point of view, that *tablet* was used by patients more than by healthy participants. Nevertheless, the use of this technology was very low for both groups.

The purpose of *socialization* should be treated carefully. In fact, *smartphone* was used for *socialization* by a high number of patients and healthy participants, although healthy subjects prevail and there is statistical significance. However, on the contrary, *game console* was much more frequently used by patients than healthy participants.

For *shopping*, the difference in the use of *computer* is statistically significant in favor of healthy participants. Moreover, all other devices were found to be less used by patients for this purpose.

As noted before, both patients and healthy participants used most devices for the same purpose. Thus, the main purpose for using *smartphone* in both groups was *socialization*, followed by *entertainment*. Similarly, both *game console* and *smart TV* were used in both groups for the main purpose of *entertainment*. The main purpose of use of *tablet* was *entertainment* for both groups.

Regarding the purpose for using the *computer*, it is noticeable that patients used it preferentially for playing compared with healthy participants who primarily used it as a work tool.

Use of the Internet

Regardless of device, what type of website do patients and healthy participants use when they access the internet? Do they use the internet to search for information or socialize? Participants were asked to indicate a maximum of 3 search engines.

Table 9 shows the outcomes regarding the search engines used by participants. The search engines were classified into *general* (general purpose), *entertainment*, *knowledge* (knowledge sharing), *shopping*, and *work*. As can be seen, the percentages are quite similar for patients and healthy participants.

As expected, a great majority of participants use *Google* as a search engine and *YouTube* for entertainment purposes and the use of specific websites for work and shopping is very low. For *work* and *shopping*, the specific search engines were not included, as almost every participant uses a different website. Moreover, the difference in the use of Wikipedia is interesting: 11.1% (10 out of 90 participants) for patients and 3.3% for healthy participants (3 out of 90 participants). When analyzing the number of answers provided by participants, only 5.6% (5 out of 90 participants) of patients did not indicate any search engine, whereas all healthy participants indicated at least one search engine. In addition, the percentage of participants indicating 3 apps was higher for healthy participants than patients.

Table 10 analyzes the apps used for socialization. As in the previous study, a classification was established to choose the main purpose of each app used, although it was difficult to categorize them into only one class.

Use of Videogames

For videogames, Table 11 demonstrates that the participants were familiar with a great number of applications. Indeed, no single videogame was the most used by participants. The highest figures were for FIFA 17 in patients and for League of Legends in healthy participants.

Table 9. Use of search engines by patients with early-stage schizophrenia and healthy participants.

Participants	Patients		Healthy participants	
	Responses (N=163), n (%)	Individuals (N=90), n (%)	Responses (N=173), n (%)	Individuals (N=90), n (%)
General	96 (59)	—[a]	106 (61)	—
Google	—	86 (96)	—	89 (99)
Yahoo	—	9 (10)	—	9 (10)
Bing	—	1 (1)	—	8 (9)
Entertainment	49 (30)	—	51 (29)	—
YouTube	—	49 (54)	—	45 (50)
Spotify	—	—	—	6 (7)
Knowledge	10 (6)	—	3 (2)	—
Wikipedia	—	10 (11)	—	3 (3)
Shopping	5 (3)	5 (6)	6 (3)	6 (7)
Work	3 (2)	3 (3)	7 (4)	7 (8)

[a] Not applicable.

Table 10. App used for socialization by patients with early-stage schizophrenia and healthy participants.

Participants	Patients		Healthy participants	
	Responses (N=181), n (%)	Individuals (N=90), n (%)	Responses (N=223), n (%)	Individuals (N=90), n (%)
Messaging	95 (52)	—[a]	118 (539)	—
WhatsApp	—	65 (72)	—	72 (80)
Skype	—	12 (13)	—	9 (10)
Twitter	—	9 (10)	—	33 (37)
Snapchat	—	6 (7)	—	1 (1)
WeChat	—	2 (2)	—	0 (0)
Google Hangouts	—	1 (1)	—	0 (0)
Telegram	—	0 (0)	—	3 (3)
Social media	53 (29)	—	65 (29)	—
Facebook	—	53 (59)	—	65 (72)
Photo sharing	28 (15)	—	35 (16)	—
Instagram	—	28 (31)	—	35 (39)
Business and employment	3 (2)	—	3 (1)	6 (7)
Catalog of ideas	1 (1)	—	1 (0)	7 (8)
Dating	1 (1)	—	1 (0)	7 (8)

[a] Not applicable.

Table 11. Use of videogames in patients with early-stage schizophrenia versus a control sample.

Videogames	Patients (N=75), n (%)	Healthy participants (N=95), n (%)
FIFA 17	12 (16)	7 (7)
League of Legends	5 (7)	10 (10)
Overwatch	2 (3)	8 (8)
Mario Kart 8	7 (9)	1 (1)
Grand Theft Auto V	6 (8)	1 (1)
The Elder Scrolls V: Skyrim	0 (0)	6 (6)
Call of Duty: Infinite Warfare	5 (7)	3 (3)
The Legend of Zelda: Breath of the Wild	1 (1)	5 (5)
Minecraft	4 (5)	2 (2)
Pokémon GO	0 (0)	4 (4)
Hearthstone: Heroes of Warcraft	0 (0)	3 (3)
The Sims 4	1 (1)	3 (3)
Others	36 (48)	42 (44)

Discussion

Principal Findings

To our knowledge, this is the first multicentric study on the use of digital technology for internet access in patients with early-stage schizophrenia compared with a sample of control participants matched in age, gender, and place of residence. The smartphone is clearly the digital device most widely used by patients (84 out of 90 participants), reaching almost 95% of use at least twice a week. The second most used device is the computer, which reaches a use of almost 75% (67 out of 90 participants) at least twice per week. With regard to the other devices, the low usage indicated and the high number of *never, but would like to* responses is striking. This might be because these devices are not economically affordable for this group of the Spanish population despite there being outstanding interest in them.

In any event, although the use of all devices is lower in patients compared with healthy participants, the results indicate that a large proportion of patients with early-stage schizophrenia have access to different technological devices and use them frequently. This finding is of great importance for designing intervention programs including the use of technological devices, especially smartphones and computers. The inclusion of technological devices is being investigated in several areas such as neurocognitive remediation [34], adherence to pharmacological treatment [35], social cognition remediation [36], treatment of refractory auditory hallucinations [37], or training in social skills [38].

Moreover, no statistically significant differences were found in the *frequency of* use of digital technology devices between male and female patients, which is interesting in terms of gender equality. In addition, no statistically significant differences were detected in terms of age, except for a lower use of smart TV in older adults. It was also found that frequency of use of computers increased significantly as educational level increased. Although not statistically significant, patients who lived in cities used most digital devices more frequently than those who lived in small towns. To sum up, healthy participants used all digital technology devices more frequently than patients, with this difference between the groups being statistically significant for all devices except for game console.

The main results in terms of *purpose of use of digital technology* when accessing the internet are described below. Patients and healthy participants coincided in that the main purpose of use for tablet, game console, and smart TV was entertainment and for smartphone, socialization. However, they differ in the purpose for using the computer, with entertainment being the principal motivation for patients and work for healthy participants. In general, most devices are less used by patients for this purpose. These data could indicate a lower interest of patients for this purpose.

Specifically, entertainment seems to be the main objective sought with most devices, both in patients and in healthy participants. For entertainment, patients use smart TV and smartphone more frequently, whereas healthy participants prefer the game console, with this difference being statistically significant. For work, there is a statistically significant difference in favor of healthy participants with respect to computer use. The devices most used to socialize are smartphone (with a statistically significant difference in favor of healthy participants) and computer in both groups. Another finding is that patients used the game console more frequently than healthy participants for socialization purposes. Web-based shopping seems to be the least common objective in both groups. Nevertheless, healthy participants engaged more in this type of activity, especially on the computer.

With regard to the most used *search tools*, the main finding is the lack of statistically significant differences between the 2 groups. Google is the most important search engine followed by YouTube. The access to social networks is quite similar in both groups, with WhatsApp and Facebook being the most important applications. As expected, the most frequently used domain of social activity was messaging, followed by social media and photo sharing. However, the use of Twitter, another well-known app in this domain, is slightly curious. Here, patients presented a lower use of this app than healthy participants, which was nevertheless mitigated through other messaging apps such as Snapchat and WeChat.

Regarding the number of answers provided by the participants, it is surprising that several patients with early-stage schizophrenia indicated no app, whereas this percentage was null for the other group. Moreover, the percentage of participants reporting a maximum of 3 apps was higher in the case of healthy participants.

In the case of *videogames*, both groups knew and played several video games, with FIFA 17 being the favorite for patients and League of Legends for healthy participants. Nonetheless, no single videogame was used by most participants. This means that an accurate analysis cannot be performed by just looking at the results provided in this table. Therefore, a deeper analysis, probably mining the most important features of the videogames, is necessary to give any concluding remarks in this regard.

Strengths and Limitations

There are several strengths and limitations in this study.

The main strength of this work is that it is the first multicentric study carried out on internet access through digital technology (with this degree of depth in terms of frequency of use, purpose of use, and type of digital technology) by patients with early-stage schizophrenia in Spain compared with a sample of controls matched in age, gender, and place of residence.

It should also be highlighted that the selection of a multicentric approach representing 3 typical, albeit different, cities is essential to generalize the results to the overall Spanish population. Indeed, this research study involved participants from 3 different recruitment centers (cities): *Hospital Virgen de la Luz* (Cuenca), *Hospital Universitario de San Juan* (Alicante), and *Hospital Universitario 12 de Octubre* (Madrid). The 3 centers of recruitment are related to 3 different types of settlements based on their number of inhabitants. The first could be classified as a *large town*, the second a *large city*, and the last a *metropolis*. Moreover, in terms of rural population, Cuenca has a high percentage, in Alicante the percentage is small, and in Madrid it is practically null.

In addition, regarding the statistical analysis, the size of the sample (180 participants) can be considered more than enough to extract important and useful considerations. In this sense, another strength of the study is having a control group of the same sociocultural environment, matched in age and gender.

There are some limitations to our study, mainly that socioeconomic status, ethnicity, culture, and working status of participants were not collected, which would have helped enhance the explanation of some results obtained.

Comparison With Prior Works

To the best of our knowledge, no other published multicentric studies have analyzed the access and use of digital technology devices and applications in patients with early-stage schizophrenia in comparison with a sample of healthy participants matched in age, gender, and place of residence.

The results of this study indicate that a large proportion of patients have access to different technological devices and use them frequently, although the use of all devices is lower in patients compared with healthy participants. This lower use could be attributed to lower economical level of the patients [39], lower academic level [40], their cognitive dysfunction [41], unemployment rates [42], or the presence of negative symptoms [43]. Our results for access to different technological devices and use coincide with the results of a previous study, which concluded that more than 90% of the psychotic patients (not only the ones in early stage) surveyed owned more than one digital device and most of them used multiple devices habitually [12].

Both patients and healthy participants coincide in the preference for using smartphone over the other devices. The second most used device is the computer, followed by smart TV, game console, and tablet. The preference for the use of smartphone conforms to the Spanish population's access to this device, as well as the preference of its use to access the internet [1]. Considering this previous report, the smartphone seems to be the most used device by the Spanish population, both for healthy participants and patients, which coincides with the results obtained in our own study.

A recent meta-analysis on smartphone ownership in psychotic patients revealed that it was increasing rapidly, with 81.4% ownership among respondents between 2014 and 2015 in the United States [44]. Another study carried out in Georgia in 2015, with a sample of 80 patients with schizophrenia, concluded that there was greater access to smartphones compared with computers in the study sample (73% vs 54%), which would also be compatible with our results [24].

In addition, in a previous survey of access and use of technological tools carried out in 2012 in Montreal on 71 patients with a first psychotic episode [29], a greater use of computer (desktop or laptop) was reported in comparison with mobile phone or smartphone, reaching 96% versus 70% of frequency (defined as *from daily use up to 2 to 3 times per week*). This difference in the preference for computer compared with smartphone could be related to 2 circumstances. The first could be cultural difference, as in 2017, the device most frequently used to access the internet was smartphone (76%), followed closely by laptop computer (71%), according to the National Statistics Institute of Canada [45]. As can be concluded from these results, the Canadian population has no marked preference for the smartphone, with both having practically the same frequency of use. The second reason could be due to the fact that the study was carried out in 2012 and the pattern of use of digital technology is constantly changing and adapting to the needs of the population. Unlike other previously published studies, a greater use of game consoles in men was not observed [29].

Some years ago, an interesting study found that the internet is an influential source of illness-related information for patients with schizophrenia [46]. Moreover, the same report stated that many aspects of their behavior related to internet usage resembled those of individuals without mental illness. The following paragraphs support this statement.

Regarding the most visited search engines, the responses of both groups (patients and healthy participants) were quite similar. Google is the most named search engine by far (with close to 100% use in both groups), which shows the enormous expansion of this tool. The results (although not as overwhelming) were similar for YouTube, the main page visited by both groups for entertainment. It is striking that both patients and healthy participants named these search tools freely, that is, without a list provided by the interviewer. At the time of data collection, both groups coincided in the use of Wikipedia, the use being higher in patients than in healthy participants. With regard to other types of websites related to job searching, the response rate was low in both groups, although higher in the healthy participants, which could be related to the lower employment rates in the psychotic population.

With regard to the applications used for socialization, it is striking that around 12% of patients use no application. This is probably related to the difficulties these patients present when socializing. Nevertheless, the patients who responded did so similarly to the healthy participants, with WhatsApp being the favorite app, followed by Facebook and Instagram. As with Google and YouTube, it can be concluded that these applications have expanded enormously in recent years, reaching even young psychotic patients. Both groups agree on the low rate of responses in the use of dating websites, which could be related to inhibition when verbalizing this type of information.

Conclusions

This is the first multicentric study carried out that provides information about the access to the internet and use of digital technology devices and applications by patients with early-stage schizophrenia in comparison with healthy participants matched in age, gender, and place of residence.

In general terms, the results obtained in our study indicated that a large proportion of patients with early-stage schizophrenia have access to different digital devices and use them frequently. However, a lower frequency of use of most devices was found in patients compared with healthy participants. For some devices, this was due to a lack of access, not an absence of interest. Nevertheless, both groups coincide in the most used devices. In addition, the purpose of using the devices in relation to the internet is highly similar in both groups.

This study brings the scientific community closer to the patterns of internet access and use of digital technology in patients with early-stage schizophrenia compared with healthy participants from the same demographic areas. The analysis of this information will be useful to guide the future development of internet technology–based therapeutic applications [47].

Acknowledgments

This work was partially supported by Spanish Instituto de Salud Carlos III/European Regional Development Fund (FEDER, UE) under PI16/00359 grant and Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI)/FEDER, UE under DPI2016-80894-R and TIN2015-72931-EXP grants, by Madrid Regional Government (R&D activities in Biomedicine S2017/BMD-3740 (AGES-CM 2-CM)) and Structural Funds of the European Union, and by the Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM) of the Instituto de Salud Carlos III. These funding sources have not been involved in study design; collection, analysis, and interpretation of data; writing of the report; or in the decision to submit the paper for publication.

Authors' Contributions

RRJ, PFS, AFC, PG, and JLS designed the study. PG, AIA, IT, MD, JLS, IMG, and LFG managed the literature searches and analyses. PFS, AIA, PFS, IMG, and LFG selected the sample, interviewed participants, and contributed in some aspects of the study design and in the interpretation of results. AFC, IT, MDT, and PG undertook the statistical analysis. PFS, AFC, PG, and RRJ wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

RRJ has been a consultant for, spoken in activities of, or received grants from Instituto de Salud Carlos III, Fondo de Investigación Sanitaria (FIS), Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid Regional Government (S2010/BMD-2422 AGES), Janssen-Cilag, Lundbeck, Otsuka, Pfizer, Ferrer, Juste, and Takeda.

Multimedia Appendix 1

Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist.

[DOCX File, 46KB - jmir_v21i4e11824_app1.docx]

References

1. Broadband Commission for Sustainable Development. International Telecommunications Union. Geneva, Switzerland: ITU/UNESCO; 2017 Sep 1. The state of broadband 2017: broadband catalyzing sustainable development URL: https://www.itu.int/dms_pub/itu-s/opb/pol/s-pol-broadband.18-2017-pdf-e.pdf [accessed 2019-02-12] [WebCite Cache ID 7699Kbjfr]
2. Instituto Nacional de Estadística. 2018 Nov 7. Survey on equipment and use of information and communication technologies in households URL: http://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608 [WebCite Cache ID 73MD1VMwV]
3. Harvey PD, Se Keefe R. Technology, society, and mental illness: challenges and opportunities for assessment and treatment. Innov Clin Neurosci 2012 Nov;9(11-12):47-50 [FREE Full text] [Medline: 23346519]
4. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265 [FREE Full text] [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
5. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Dec;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
6. Silva BM, Rodrigues JJ, de la Torre DI, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
7. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
8. Evans WD, Abroms LC, Poropatich R, Nielsen PE, Wallace JL. Mobile health evaluation methods: the Text4baby case study. J Health Commun 2012 May;17(Suppl 1):22-29. [doi: 10.1080/10810730.2011.649157] [Medline: 22548595]
9. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
10. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
11. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using Internet-and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017 Aug;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
12. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 4;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]

13. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
14. Ashford RD, Lynch K, Curtis B. Technology and social media use among patients enrolled in outpatient addiction treatment programs: cross-sectional survey study. J Med Internet Res 2018 Mar 6;20(3):e84 [FREE Full text] [doi: 10.2196/jmir.9172] [Medline: 29510968]
15. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Dec 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
16. 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]
17. 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;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
18. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
19. Patel A, Everitt B, Knapp M, Reeder C, Grant D, Ecker C, et al. Schizophrenia patients with cognitive deficits: factors associated with costs. Schizophr Bull 2006 Oct;32(4):776-785 [FREE Full text] [doi: 10.1093/schbul/sbl013] [Medline: 16885205]
20. van Os J, Kapur S. Schizophrenia. Lancet 2009 Aug 22;374(9690):635-645. [doi: 10.1016/S0140-6736(09)60995-8] [Medline: 19700006]
21. Gire N, Farooq S, Naeem F, Duxbury J, McKeown M, Kundi PS, et al. mHealth based interventions for the assessment and treatment of psychotic disorders: a systematic review. Mhealth 2017;3:33 [FREE Full text] [doi: 10.21037/mhealth.2017.07.03] [Medline: 28894743]
22. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 May;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
23. Birnbaum ML, Ernala SK, Rizvi AF, de Choudhury CM, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Aug 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
24. Miller BJ, Stewart A, Schrimsher J, Peeples D, Buckley PF. How connected are people with schizophrenia? Cell phone, computer, email, and social media use. Psychiatry Res 2015 Feb 28;225(3):458-463. [doi: 10.1016/j.psychres.2014.11.067] [Medline: 25563669]
25. Brunault P, Bray A, Rerolle C, Cognet S, Gaillard P, El-Hage W. [Differences in access to internet and internet-based information seeking according to the type of psychiatric disorder]. Rev Epidemiol Sante Publique 2017 Apr;65(2):125-136. [doi: 10.1016/j.respe.2016.11.001] [Medline: 28153646]
26. Thomas N, Foley F, Lindblom K, Lee S. Are people with severe mental illness ready for online interventions? Access and use of the internet in Australian mental health service users. Australas Psychiatry 2017 Jun;25(3):257-261. [doi: 10.1177/1039856217689913] [Medline: 28139947]
27. Athanasopoulou C, Välimäki M, Koutra K, Löttyniemi E, Bertsias A, Basta M, et al. Internet use, eHealth literacy and attitudes toward computer/internet among people with schizophrenia spectrum disorders: a cross-sectional study in two distant European regions. BMC Med Inform Decis Mak 2017 Sep 20;17(1):136 [FREE Full text] [doi: 10.1186/s12911-017-0531-4] [Medline: 28931385]
28. Välimäki M, Kuosmanen L, Hätönen H, Koivunen M, Pitkänen A, Athanasopoulou C, et al. Connectivity to computers and the Internet among patients with schizophrenia spectrum disorders: a cross-sectional study. Neuropsychiatr Dis Treat 2017 Apr;13:1201-1209 [FREE Full text] [doi: 10.2147/NDT.S130818] [Medline: 28490882]
29. Abdel-Baki A, Lal S, D-Charron O, Stip E, Kara N. Understanding access and use of technology among youth with first-episode psychosis to inform the development of technology-enabled therapeutic interventions. Early Interv Psychiatry 2017 Dec;11(1):72-76. [doi: 10.1111/eip.12250] [Medline: 26011657]
30. Bonet L, Llácer B, Hernandez-Viadel M, Arce D, Blanquer I, Cañete C, et al. Differences in the use and opinions about new eHealth technologies among patients with psychosis: structured questionnaire. JMIR Ment Health 2018 Jul 25;5(3):e51 [FREE Full text] [doi: 10.2196/mental.9950] [Medline: 30045835]
31. First M, Williams J, Karg R, Spitzer R. Structured Clinical Interview for DSM-5, Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA: American Psychiatric Association; 2015:1-94.
32. Sánchez-Morla EM, Barabash A, Martínez-Vizcaíno V, Tabarés-Seisdedos R, Balanzá-Martínez V, Cabranes-Díaz JA, et al. Comparative study of neurocognitive function in euthymic bipolar patients and stabilized schizophrenic patients. Psychiatry Research 2009 Oct;169(3):220-228. [doi: 10.1016/j.psychres.2008.06.032] [Medline: 19758705]

33. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
34. Garrido G, Penadés R, Barrios M, Aragay N, Ramos I, Vallès V, et al. Computer-assisted cognitive remediation therapy in schizophrenia: durability of the effects and cost-utility analysis. Psychiatry Res 2017 Dec;254:198-204. [doi: 10.1016/j.psychres.2017.04.065] [Medline: 28463718]
35. Batra S, Baker R, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
36. Fernandez-Gonzalo S, Turon M, Jodar M, Pousa E, Hernandez Rambla C, García R, et al. A new computerized cognitive and social cognition training specifically designed for patients with schizophrenia/schizoaffective disorder in early stages of illness: A pilot study. Psychiatry Research 2015 Aug;228(3):501-509. [doi: 10.1016/j.psychres.2015.06.007] [Medline: 26163731]
37. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. The Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/S2215-0366(17)30427-3]
38. Rus-Calafell M, Gutiérrez-Maldonado J, Ribas-Sabaté J. A virtual reality-integrated program for improving social skills in patients with schizophrenia: a pilot study. J Behav Ther Exp Psychiatry 2014 Mar;45(1):81-89. [doi: 10.1016/j.jbtep.2013.09.002] [Medline: 24063993]
39. Werner S, Malaspina D, Rabinowitz J. Socioeconomic status at birth is associated with risk of schizophrenia: population-based multilevel study. Schizophr Bull 2007 Nov;33(6):1373-1378 [FREE Full text] [doi: 10.1093/schbul/sbm032] [Medline: 17443013]
40. Swanson C, Gur RC, Bilker W, Petty RG, Gur RE. Premorbid educational attainment in schizophrenia: association with symptoms, functioning, and neurobehavioral measures. Biol Psychiatry 1998 Oct 15;44(8):739-747. [doi: 10.1016/S0006-3223(98)00046-8] [Medline: 9798078]
41. Fett AJ, Viechtbauer W, Dominguez M, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci Biobehav Rev 2011 Jan;35(3):573-588. [doi: 10.1016/j.neubiorev.2010.07.001] [Medline: 20620163]
42. 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]
43. Kirkpatrick B, Fenton WS, Carpenter WT, Marder SR. The NIMH-MATRICS consensus statement on negative symptoms. Schizophr Bull 2006 Apr;32(2):214-219 [FREE Full text] [doi: 10.1093/schbul/sbj053] [Medline: 16481659]
44. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 6;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
45. Statistics Canada. 2016. General Social Survey: Canadians at Work and Home (GSS) URL: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5221 [WebCite Cache ID 73MELyFsu]
46. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
47. Fernández-Caballero A, Navarro E, Fernández-Sotos P, González P, Ricarte JJ, Latorre JM, et al. Human-avatar symbiosis for the treatment of auditory verbal hallucinations in schizophrenia through virtual/augmented reality and brain-computer interfaces. Front Neuroinform 2017 Nov;11:64 [FREE Full text] [doi: 10.3389/fninf.2017.00064] [Medline: 29209193]

Abbreviations

eHealth: electronic health
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, fifth edition
INE: Instituto Nacional de Estadística
ITU: International Telecommunication Union
mHealth: mobile health

Edited by G Eysenbach; submitted 03.08.18; peer-reviewed by F Holl, K Mishina, A Daros; comments to author 12.10.18; revised version received 03.12.18; accepted 12.12.18; published 05.04.19.

Please cite as:
Fernández-Sotos P, Fernández-Caballero A, González P, Aparicio AI, Martínez-Gras I, Torio I, Dompablo M, García-Fernández L, Santos JL, Rodriguez-Jimenez R
Digital Technology for Internet Access by Patients With Early-Stage Schizophrenia in Spain: Multicenter Research Study
J Med Internet Res 2019;21(4):e11824
URL: https://www.jmir.org/2019/4/e11824/
doi:10.2196/11824
PMID:30950798

Review

Internet-Delivered Cognitive Behavioral Therapy for Anxiety Disorders in Open Community Versus Clinical Service Recruitment: Meta-Analysis

Geke Romijn[1,2], MSc; Neeltje Batelaan[3], MD, PhD; Robin Kok[1,4,5], PhD; Jeroen Koning[6], MD, PhD; Anton van Balkom[3], MD, PhD; Nickolai Titov[7], PhD; Heleen Riper[1,8], PhD

[1]Department of Clinical, Neuro and Developmental Psychology, VU University, Amsterdam Public Health Research Institute, Amsterdam, Netherlands
[2]Altrecht Academic Anxiety Centre, Utrecht, Netherlands
[3]Department of Psychiatry, VU University Medical Centre, Amsterdam Public Health Research Institute, Amsterdam, Netherlands
[4]Department of Psychology, University of Southern Denmark, Odense, Denmark
[5]Centre for Innovative Medical Technology, Odense University Hospital, Odense, Denmark
[6]Pro Persona Psychiatric Centre, Tiel, Netherlands
[7]Mindspot, Department of Psychology, Macquarie University, Sydney, Australia
[8]Centre for Telepsychiatry, Mental Health Services of Southern Denmark, Odense, Denmark

Corresponding Author:
Geke Romijn, MSc
Department of Clinical, Neuro and Developmental Psychology
VU University
Amsterdam Public Health Research Institute
Van der Boechorstraat 7
Amsterdam,
Netherlands
Phone: 31 20 59 89009
Fax: 31 20 59 88758
Email: g.a.romijn@vu.nl

Abstract

Background: Ample studies have shown the effectiveness of internet-delivered cognitive behavioral therapy (iCBT) for anxiety disorders. These studies recruited their participants mainly from the community and, to a lesser extent, from within routine care services. Little is known about whether different recruitment strategies lead to different treatment effects.

Objective: This meta-analysis compared clinical results obtained in trials with recruitment from the community versus results obtained in trials with clinical service recruitment and explored factors that may mediate differences in treatment outcome.

Methods: We included randomized controlled trials in which the clinical effects of iCBT for anxiety disorders were compared with a control condition (waitlist controls or face-to-face cognitive behavioral therapy). We classified trials as open recruitment trials (recruitment from the community) or clinical service recruitment trials (recruitment through outpatient clinics). Pooled effect sizes based on measures examining anxiety symptoms, depressive symptoms, and quality of life were computed for each type of trial. Subgroup analyses examined whether clinical results from open recruitment trials differed from those obtained in clinical service recruitment trials. Additional analyses explored which demographic, clinical, and treatment-related factors contributed to differences in effect sizes of open recruitment versus clinical service recruitment trials.

Results: We included 42 studies with 53 comparisons (43 open recruitment comparisons and 10 clinical recruitment comparisons). Analyses of anxiety measures revealed, first, that iCBT open recruitment studies with waitlist control comparators showed a significantly higher effect size for decrease in anxiety symptoms than did those with clinical recruitment ($Q=10.09$; $P=.001$). This association between recruitment method and effect size was no longer significant in a multivariate metaregression with treatment adherence and exclusion of patients with depressive symptoms entered as additional predictors of effect size. Second, effect size for decrease in anxiety symptoms did not differ significantly between clinical recruitment and open recruitment studies with face-to-face cognitive behavioral therapy comparators. The effects of open recruitment trials and clinical recruitment trials did not differ significantly for the secondary outcomes, compared with face-to-face cognitive behavioral therapy and waitlist controls.

Conclusions: iCBT was effective in samples recruited in clinical practice, but effect sizes were smaller than those found in trials with an open recruitment method for studies with waitlist control comparators. Hence, for patients with anxiety disorders in routine care, the impact of iCBT may not be as positive as for study participants recruited from the community. The difference between open recruitment trials and clinical service recruitment trials might be partly explained by patients' greater therapy adherence in open recruitment trials and the stricter exclusion of patients with severe depressive symptoms in these studies. Since most trials in this meta-analysis applied an open recruitment method, more studies with routine care populations are needed to further validate these findings.

(J Med Internet Res 2019;21(4):e11706) doi:10.2196/11706

KEYWORDS

anxiety disorders; cognitive behavioral therapy; internet; recruitment method; efficacy; effectiveness

Introduction

Background

Internet-delivered cognitive behavioral therapy (iCBT) for anxiety disorders has been tested in ample randomized controlled trials and several meta-analyses. These studies show the potential of iCBT to reduce anxiety symptoms among patients and general populations, indicating that iCBT is effective when compared with a waitlist control (WLC), with effect sizes in the moderate to large range. Studies also suggest that it is as effective as face-to-face cognitive behavioral therapy (CBT) in improving symptoms of anxiety [1-5], although these studies are limited in number. Furthermore, iCBT may minimize treatment barriers such as high costs due to reduced time needed by therapists to provide therapy [6,7] and scalability.

The majority of trials on iCBT apply an open recruitment (OR) strategy, inviting individuals with anxiety symptoms from within the community to directly partake in the research study. These participants refer themselves to such a study. Often these studies apply strict inclusion and exclusion criteria; for example, they may exclude patients taking psychoactive medication, patients with comorbid disorders, or severely depressed patients [8,9]. Clinical service recruitment (CSR) trials, on the other hand, invite patients already seeking treatment in clinical practices to participate. Trials with an OR method provide evidence more related to *efficacy* (investigating whether a treatment works under ideal circumstances, with high internal validity), as opposed to CSR trials that are more related to *effectiveness* and provide information on whether a treatment works in clinically representative conditions [10].

Only a minority of the patient samples in trials with an OR method correspond to patient populations in a regular clinical setting in terms of sociodemographic characteristics, motivation for treatment, level of suffering, and clinical characteristics such as severity of anxiety, comorbidity, or medical history [11-13]. Furthermore, the use of extensive exclusion criteria in OR trials can reduce the degree to which these study samples resemble clinical populations in routine care settings. A meta-analysis [14] found a strong and positive relationship (r=.70) between the number of exclusion criteria and the rate of clinically improved participants for studies on iCBT for anxiety disorders. These results suggest a lower clinical effectiveness in clinically representative studies than in highly controlled studies. This raises the question whether results from OR trials can be extrapolated to routine clinical practice.

On the other hand, uncontrolled effectiveness studies show large clinical effects [15-20], thereby suggesting that iCBT for anxiety disorders may be as effective in routine care settings as demonstrated in efficacy trials. One review investigated controlled research of iCBT in routine clinical practice [21]. Results showed that effect sizes obtained from effectiveness studies (ranging from 0.75 to 1.73) were in the same range as those obtained in efficacy trials, though only 3 randomized controlled trials were included.

Objective

Although several meta-analyses for (internet-based treatment of) anxiety disorders have been conducted in recent years, to our knowledge, none of these studies have compared the potential differences in clinical effectiveness between OR and CSR trials. In this study, we aimed to (1) assess whether OR trials produced clinical effectiveness for anxiety symptoms similar to that of CSR trials and (2) explore predictors of potential effect differences, such as demographic, clinical, and treatment-related characteristics. We based these predictors on differences between OR and CSR trials in patient samples and methods found in previous studies [11,13,14].

Methods

Study Retrieval

We report this meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [22]. We retrieved studies through systematic literature searches in PubMed, PsycINFO, and EMBASE databases. Searches were conducted with keywords and text words, in which words indicative of internet treatment were combined with words indicative of anxiety disorder, CBT, and randomized controlled trial (see Multimedia Appendix 1 for the full search string). Furthermore, we checked reference lists of retrieved articles and of earlier reviews on iCBT for anxiety disorders [1-4].

Inclusion and Exclusion Criteria

We included randomized controlled trials published up to and including December 2017 on guided and unguided iCBT and blended CBT for adults. Blended CBT combines face-to-face treatment with internet components into a single integrated treatment protocol [23]. We included only randomized controlled trials that assessed a primary diagnosis of an anxiety disorder according to the *Diagnostic and Statistical Manual of Mental Disorders* (Third Edition Revised [DSM-III-R], DSM (Fourth Edition [DSM-IV]), or DSM (Fifth Edition [DSM-5]) established by a structured diagnostic interview. We excluded studies on obsessive compulsive disorder [24] and posttraumatic stress disorder [25], since they are not classified as anxiety disorders in DSM-5. We included only randomized controlled trials published in English or Dutch.

Interventions and Comparators

We considered interventions to be CBT if they were based on cognitive behavioral principles [26] and consisted at least of cognitive restructuring or exposure (interoceptive exposure or exposure in vivo), or a combination of both. To be considered iCBT, the intervention must have been delivered (partly) via a computer or the internet through the use of webpages or email, or both. We included studies on iCBT targeting anxiety disorders and studies on transdiagnostic iCBT [27-31], addressing multiple anxiety disorders or addressing both anxiety and mood disorders, but only if participants had a diagnosis of an anxiety disorder and measures of anxiety were reported. We did not include interventions when the Web-based part of the treatment was limited to exposure scenes on a screen (eg, Heading et al [32]) because we considered this to be in virtuo exposure treatment, which is beyond the scope of this review.

iCBT was compared with WLC or regular face-to-face CBT treatment (including individual or group CBT delivered in a face-to-face format). We excluded studies with other comparisons such as transdiagnostic iCBT compared with disorder-specific iCBT, or guided iCBT versus self-help [24,25,33,34].

Outcome Measures

Our primary outcome was anxiety symptom severity based on the score on a rating scale used to measure general symptoms of anxiety. We applied a hierarchy of preferred outcomes for all measures based on frequency of use in the included trials. For general measures of anxiety, the preferred order was as follows: Beck Anxiety Inventory (BAI [35]), anxiety scale of the Depression Anxiety Stress Scales [36], State-Trait Anxiety Inventory [37], and Anxiety Sensitivity Index [38]. When a general measure of anxiety was not available, we used a measure for specific anxiety symptoms (see Multimedia Appendix 2 [39-49] for the order of rating scales and Multimedia Appendix 3 [7,27-31,50-85] for the outcome measures we used for all studies).

Since anxiety disorders are frequently accompanied by symptoms of depression and a reduced quality of life [86], secondary outcome measures were effects on depression severity and on quality of life (see Multimedia Appendix 2 for the order of rating scales [36,87-94]).

Criteria for Open Recruitment and Clinical Service Recruitment

We classified trials as OR trials if participants were recruited from the community and referred themselves to be interested in the study in response to the invitation from a research team, by means of advertisements in newspapers or magazines, banners on websites, or large-scale mailings. In CSR trials, recruitment was carried out among patients already seeking treatment in outpatient clinical mental health practices. In the case of mixed recruitment strategies, we classified trials according to the most prominent recruitment strategy.

Study Selection and Data Extraction

Two of the authors (GR and NB) independently screened the list of titles and abstracts that resulted from the literature search. Reference lists were screened for additional studies of relevance. We obtained full articles for potentially relevant abstracts according to the inclusion criteria. If included trials did not provide complete information, we contacted the primary investigator by email to attempt to obtain unreported data. We sent a second email when we received no response. Two researchers (GR and NB) extracted the data using Excel (2013) spreadsheets (Microsoft Corporation) and differences in such data were resolved by discussion. Extracted data (see Multimedia Appendix 3) included the study characteristics outlined in Textbox 1.

Textbox 1. Study characteristics extracted from the articles.

- Year of publication
- Number of participants
- Recruitment setting (open or clinical service recruitment)
- Demographic characteristics of participants included in the study (sex, age, employment status [total rate of employed participants and rate of full-time employed participants], education level [rate of participants with college degree or higher])
- Anxiety severity at baseline
- Axis I comorbidity rate
- Exclusion criteria with regard to medication use (benzodiazepines and other psychoactive medication) and depressive symptoms (indicated by a score above a cutoff level on an outcome measure for depressive symptoms)
- Details of treatment conditions (duration and type of support provided by professionals)
- Outcome data
- Number of therapy sessions according to protocol, treatment dose (number of completed therapy sessions), and treatment adherence (number of completed sessions divided by the total number of sessions according to protocol)

Risk-of-Bias Assessment

Two authors (GR and RK) independently assessed the risk of bias in the included studies based on 6 areas according to the Cochrane tool for assessing risk of bias [95]: (1) adequate generation of allocation sequence, (2) concealment of allocation to conditions (concealing allocation sequence from participants and investigators), (3) blinding of participants and personnel, (4) blinding of outcome assessors, (5) dealing with incomplete outcome data, and (6) selective outcome reporting (reported results give reason to suspect differences between reported and unreported findings). Because RK was an author of one of the included studies [50], this study was independently assessed by a third reviewer (NB). Discrepancies in scoring were resolved through discussion.

We assessed all areas as low, high, or unclear (ie, not enough information) risk of bias (see Multimedia Appendix 4). We assessed selective outcome reporting by comparing trial registrations with published articles, if available. When primary or secondary outcomes were missing, inserted, or changed in the article compared with the trial registration, or if secondary and primary outcomes had been switched, we deemed a study to be at high risk of selective outcome reporting. If no trial registration was available for a study, we coded the study as being at unknown risk for selective outcome reporting.

Statistical Analyses

We used descriptive statistics to summarize demographic characteristics, and clinical and treatment-related characteristics of OR and CSR trials. We compared categorical variables using chi-square tests and continuous variables using *t* tests. We also compared the percentage of at-risk OR and CSR trials for all risk-of-bias indicators with chi-square tests.

We then calculated the pooled overall effect sizes (Hedges *g*) indicating the difference between the conditions at posttest and their 95% confidence intervals using the random-effects model with Comprehensive Meta-Analysis software version 3.0 (Biostat). Hedges *g* is an effect size that corrects for biases due to small sample sizes [96]. Effect sizes of 0.2, 0.5, and 0.8 indicate a small, moderate, and large effect size, respectively [97]. We used the effect sizes based on intent-to-treat analysis when available (in 51 comparisons); otherwise, we used complete-sample analysis results (in 2 comparisons [51,52]).

We examined heterogeneity among studies using Higgins I^2 statistic. I^2=0% reflects no heterogeneity; 25%, 50%, and 75% indicate a low, medium, and high level of heterogeneity, respectively [98]. A higher observed statistical heterogeneity indicates a higher proportion of observed variance, which can point to underlying differences between the pooled studies. This makes interpreting the pooled effect size difficult, as it is hard to distinguish the observed effect size from the true population effect size [99]. We also calculated 95% confidence intervals around I^2 with the noncentral chi-square approach in the heterogi module for Stata 13.0SE (StataCorp LLC) [100].

We first calculated overall effect sizes on anxiety, depression, and quality-of-life treatment outcomes of iCBT compared with WLC and compared with face-to-face CBT. We then carried out subgroup analyses to assess whether clinical results from OR trials differed from those obtained in CSR trials. We performed these subgroup analyses according to the mixed-effects model, in which studies within subgroups are pooled with the random-effects model, and the fixed-effects model is used to test for significant differences among them by the between-subgroups Q-statistic [101].

We tested publication bias by inspecting the funnel plot and Egger test [102] on our primary outcome measure and by the Duval and Tweedie trim-and-fill procedure [103].

To more fully understand differences in effect size between OR and CSR trials, we conducted additional exploratory analyses. By means of subgroup and metaregression analyses, we examined which demographic, clinical, and treatment-related factors differed between OR and CSR trials and were associated with effect size on the primary outcome. Next, to examine whether these predictors contributed to the difference of effect size between OR and CSR trials, we tested recruitment method and the significant predictors in a multivariate model, except in case of collinearity. We assessed possible collinearity problems between predictors with the variance inflation factors. We considered variance inflation factor scores higher than 2.5 to indicate multicollinearity [104].

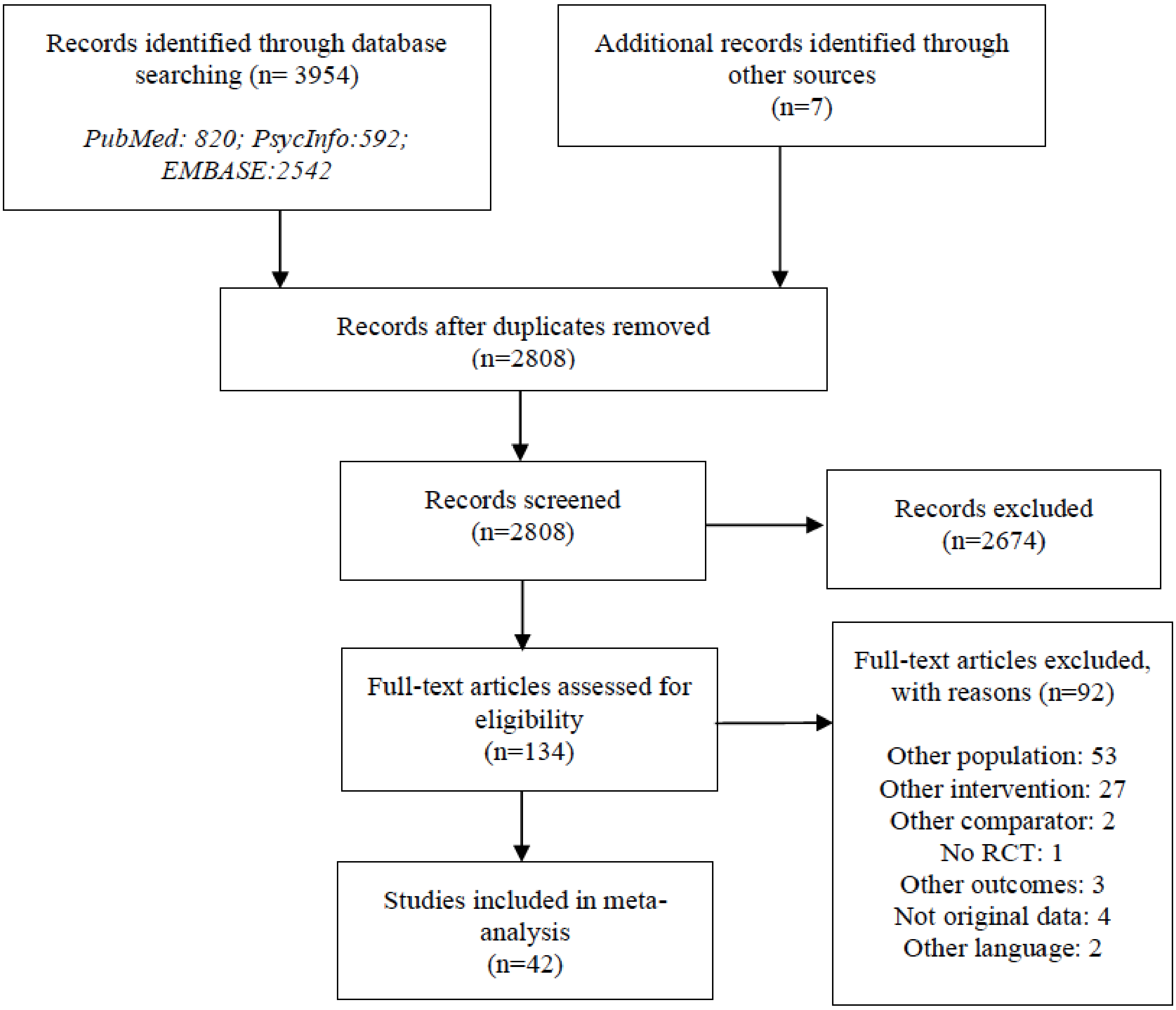
We also calculated the number needed to treat (NNT), according to Kraemer and Kupfer [105], and rounded upward to the next higher whole number [106]. The NNT gives some clinical context to statistical information, as it translates the magnitude of a statistical effect size into clinical implications—that is, the number of patients who must be treated to generate one more positive outcome than the same number of patients in the control group.

Results

Study Inclusion

The literature searches retrieved a total of 3954 abstracts. Checking references of earlier reviews resulted in 7 more citations for consideration. After we removed duplicates, we screened 2808 abstracts. After screening abstracts, we retrieved 134 full-text articles for a more detailed evaluation of eligibility. Subsequently, we excluded 92 articles because they did not meet the inclusion criteria (Figure 1). We did not include 1 study because means and standard deviations for anxiety measures were not reported [107] and we received no response from the addressed researchers to our email questions regarding these issues.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of the selection and inclusion process. RCT: randomized controlled trial.



Description of Included Studies

We included a total of 42 trials (Multimedia Appendix 3). These trials entailed 53 comparisons of anxiety outcomes between iCBT and a control condition (WLC n=41, face-to-face CBT n=12) and included 3714 participants. A total of 45 comparisons entailed outcomes on depression and 21, on quality of life. Most studies were OR trials (31 trials with 41 comparisons), versus 8 CSR trials with 9 comparisons. In 3 studies both recruitment strategies were used [31,51,53]. In 2 of these studies, most participants were self-referred (n=129, 92.8% [31] and n=70, 77% [53]) and therefore we classified these studies as OR trials. In the third study, most participants were recruited through a clinical procedure (n=76, 67%) and therefore we classified this study as a CSR trial [51]. Thus, we classified a total of 33 trials with 43 comparisons as OR trials and 9 trials with 10 comparisons as CSR trials.

iCBT typically consisted of weekly sessions (ranging from 4 to 12 sessions for studies comparing iCBT with WLC and from 4 to 23 sessions for studies comparing iCBT with face-to-race CBT), with durations ranging from 4 weeks to longer than 3 months [54]. In 4 trials a self-help iCBT intervention (unguided) was offered; in the other trials iCBT was guided, meaning that online professional support was provided.

Table 1 displays demographic, clinical, and treatment-related characteristics of OR and CSR trials. We found significant differences between OR trials and CSR trials with regard to age of participants, sex, baseline severity, exclusion of severely depressed patients, treatment dose, and treatment adherence. We could compare baseline severity only for scores based on the BAI and the Social Phobia Scale, since these were the only outcome measures that were reported in both OR trials and CSR trials. We did not classify 3 OR trials [55-57] and 1 CSR trial [51] that mentioned exclusion of patients with depressive symptoms as such, because no definition in terms of a score on a measurement for depressive symptoms was provided.

Table 1. Comparison of demographic, clinical, and treatment-related characteristics of included open recruitment studies and clinical service recruitment studies[a].

Characteristics	Open recruitment	Clinical recruitment	Test statistic	*P* value
Demographic characteristics				
Age (years), mean (SD)	37.5 (5.4)	34.1 (1.7)	t_{3657}=29.1	<.001
Female sex, n (%)	1871 (67)	541 (58.7)	χ^2_1=21.2	<.001
Education: college degree, n (%)	670 (53.4)	171 (49.4)	χ^2_1=1.7	.19
Employed: full-time or part-time, n (%)	390 (58.7)	279 (59.7)	χ^2_1=0.1	.71
Clinical characteristics				
Primary diagnosis: trials that applied this diagnosis as inclusion criterion, n (%)			χ^2_4=4.3	.37
Panic disorder	13 (30.2)	3 (30)		
Social anxiety disorder	14 (32.6)	3 (30)		
Generalized anxiety disorder	7 (16.3)	0 (0)		
Specific phobia	2 (4.7)	0 (0)		
Multiple anxiety disorders	7 (16.3)	4 (40)		
Baseline Beck Anxiety Inventory score, mean (SD)	25.4 (11.8)	29.0 (11.1)	t_{1208}=–6.2	<.001
Baseline Social Phobia Scale score, mean (SD)	36.0 (1.7)	43.0 (0.4)	t_{606}=–82.7	<.001
Comorbidity: comorbid Axis I diagnosis, %	57.6	58.1	χ^2_1=0.0	.87
Exclusion benzodiazepines: trials that applied this criterion, n (%)	10 (23.3)	0 (0)	χ^2_1=2.2	.14
Exclusion psychoactive medication other than benzodiazepines: trials that applied this criterion, n (%)	0 (0)	0 (0)	N/A[b]	N/A
Exclusion severe depression: trials that applied this criterion, n (%)	25 (58.1)	2 (20)	χ^2_1=4.7	.03
Treatment-related characteristics				
Treatment dose: number of completed sessions, mean (SD)	6.1 (1.4)	5.1 (2.4)	t_{835}=9.45	<.001
Treatment adherence: treatment completed %	77.9	53.1	t_{1599}=44.73	<.001

[a] Means and percentages are based on studies these data were available for. All available data are reported in Multimedia Appendix 3.

[b] N/A: not applicable.

No other variables differed (see Table 1). None of the included studies excluded patients who used psychoactive medication other than benzodiazepines, although a stable dose for the past 1 to 3 months was a criterion for inclusion in general.

Risk-of-Bias Assessment

For both OR trials and CSR trials, most of the studies scored a low risk on sequence allocation (OR: 27/33, 82%; CSR: 8/9, 89%), blinding of outcome assessors (OR: 33/33, 100%; CSR: 8/9, 89%), and completeness of outcome data (OR: 27/33, 82%; CSR: 6/9, 67%) (see Multimedia Appendix 4). On allocation concealment most CSR trials scored a low risk (7/9, 78%) compared with 12 of 33 (36%) OR trials. Only 4 of 33 (12%) OR and 2 of 9 (22%) CSR trials scored a low risk on selective outcome reporting, A total of 10 (30%) OR trials and 5 (56%) CSR trials scored a high risk because preregistered outcome measurements were not reported, or other outcome measurements that were not preregistered were inserted in the article. Additionally, 19 (58%) OR trials and 2 (22%) CSR trials were not registered in a trial database and we therefore scored them as having an unclear risk. We rated all included studies as having a high risk of bias on blinding of participants and personnel, because it is not possible to blind participants or therapists to the characteristics of the treatment that is offered.

We found no significant difference between the percentage of OR trials and the percentage of CSR trials with a high risk for any of the risk-of-bias indicators (*P* values ranged from *P*=.08 for sequence generation to *P*=.49 for allocation concealment).

Overall Effect

Primary Outcome

The overall mean between-groups effect size of iCBT on anxiety symptom reduction when compared with WLC at posttest was *g*=0.72 (95% CI 0.60-0.83; *P*<.001) with moderate heterogeneity of I^2=53% (95% CI 31-66) and NNT=3.

The difference in overall effect size for the decrease in anxiety symptoms between iCBT and face-to-face CBT at posttest was nonsignificant (*g*=0.12, 95% CI –0.02 to 0.26; *P*=.11; I²=0%, 95% CI 0-75; NNT=15).

Secondary Outcomes

Effect sizes of iCBT compared with WLC on depressive symptoms (*g*=0.61, 95% CI 0.46-0.75; *P*<.001; I²=70%, 95% CI 57-78; NNT=3) and quality-of-life measurements (*g*=0.44, 95% CI 0.33-0.55; *P*<.001; I²=5%, 95% CI 0-54; NNT=5) were moderate.

For iCBT compared with face-to-face CBT, effect sizes on depression measurements (*g*=0.04, 95% CI –0.13 to 0.21; *P*=.65; I²=19%. 95% CI 0-61; NNT=45) and quality-of-life outcomes (*g*=0.18, 95% CI –0.05 to 0.41; *P*=.12; I²=0%, 95% CI 0-85; NNT=10) were both nonsignificant.

Open Recruitment Versus Clinical Service Recruitment

Primary Outcome

For studies with WLC comparators, we found a significant difference between OR and CSR trials) in favor of OR trials (Q=10.09; *P*=.001) (Table 2 and Figure 2). The effect size on anxiety symptom reduction for OR trials was significant and large (*g*=0.79; *P*<.001) in favor of iCBT, whereas CSR trials obtained a small effect size (*g*=0.28; *P*=.003) in favor of iCBT.

We found no difference in anxiety symptom reduction between OR (n=6) and CSR trials (n=6) comparing iCBT with face-to-face CBT (Q=0.82; *P*=.37*)* (Table 2 and Figure 3). Both OR trials (*g*=0.19; *P*=.09) and CSR trials (*g*=0.06; *P*=.51) reported a nonsignificant difference between iCBT and face-to-face CBT on decrease in anxiety symptoms.

Secondary Outcomes

With regard to depressive symptoms, we found no significant difference between OR trials and CSR trials with WLC comparators (Q=1.43; *P*=.23) or face-to-face comparators (Q=0.85; *P*=.36).

For quality-of-life measurements, we found no significant difference between OR trials and CSR trials for studies comparing iCBT with WLC (Q=0.05; *P*=.83) or for studies comparing iCBT with face-to-face CBT (Q=0.48; *P*=.49).

Multimedia Appendix 5 presents a complete overview of results of OR and CSR subgroup analyses of secondary outcomes.

Publication Bias

Neither visual inspection of the funnel plots (see Multimedia Appendix 6) and Egger test (WLC studies: intercept=0.83; 95% CI –0.90 to 2.56; *P*=.34; face-to-face CBT studies: intercept=1.12, 95% CI –0.93 to 3.17; *P*=.25) nor the Duval and Tweedie trim-and-fill procedure showed evidence of publication bias.

Additional Exploratory Analyses

As Table 1 shows, several demographic (age, sex), clinical (baseline severity, exclusion of severely depressive patients), and treatment-related (treatment dose, treatment adherence) variables differed significantly between OR trials and CSR trials. Of these variables, only exclusion of severely depressed patients (Q=8.06; *P*=.005), treatment dose (slope=0.10; *P*=.003), and treatment adherence (slope=0.01; *P*<.001) appeared to be significantly associated with effect size for WLC comparators in separate subgroup (exclusion of severely depressed patients) and metaregression (treatment dose, treatment adherence) analyses (see Multimedia Appendix 7), meaning that the effect size was higher when severely depressed patients were excluded, when the treatment dose was higher, and when the adherence rate was higher.

In a multivariate analysis we explored whether the association between recruitment method and effect size for studies with WLC comparators was mediated by these variables. As the variance inflation factors between treatment dose and treatment adherence was 2.7, and treatment adherence was more significantly associated with effect size than treatment dose, we did not include treatment dose in the multivariate model.

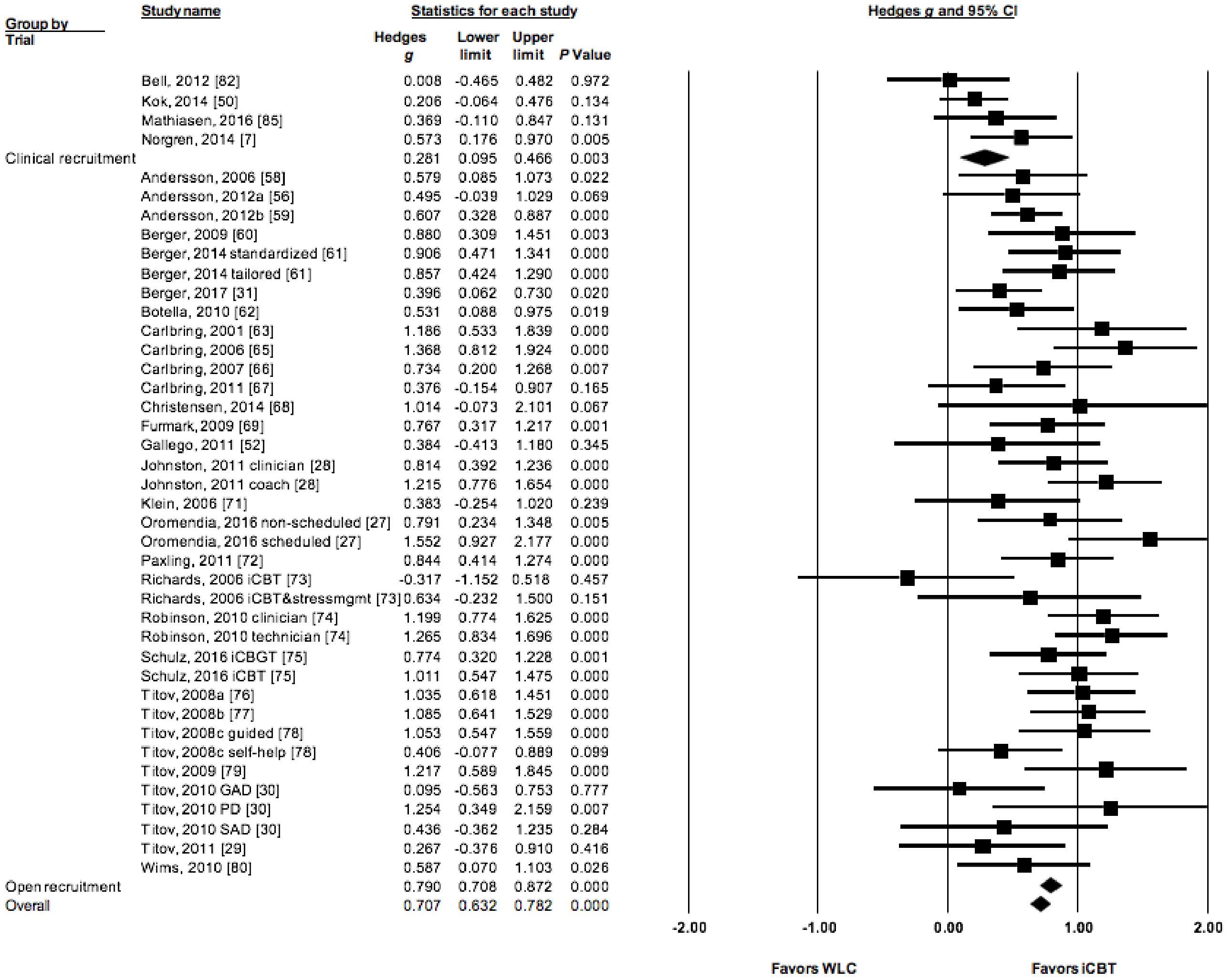
Results showed that recruitment type (slope=0.30; *P*=.14) was no longer significantly associated with the effect size in the multivariate metaregression analysis, nor was treatment adherence (slope=0.01; *P*=.23) or exclusion of severely depressed patients (slope=0.13; *P*=.27).

Table 2. Main effects of open recruitment trials and clinical service recruitment trials comparing internet-delivered cognitive behavioral therapy versus waitlist control and versus face-to-face cognitive behavioral therapy at posttest, primary outcome.

Group	Comparisons, n	Respondents, n	Hedges *g* (95% CI)	*P* value	I^2 (95% CI)	NNT[a]	Between-groups Q (*P* value)
Waitlist control							10.09 (.001)
Open recruitment	37	2474	0.79 (0.71 to 0.87)	<.001	44 (6 to 58)	3	
Clinical recruitment	4	446	0.28 (0.10 to 0.47)	.003	20 (0 to 85)	7	
Face-to-face cognitive behavioral therapy							0.82 (.37)
Open recruitment	6	336	0.19 (–0.03 to 0.40)	.09	0 (0 to 75)	10	
Clinical recruitment	6	452	0.06 (–0.12 to 0.24)	.53	0 (0 to 75)	30	

[a] NNT: number needed to treat.

Figure 2. Forest plot of effects on anxiety symptoms of open recruitment trials and clinical service recruitment trials comparing internet-delivered cognitive behavioral therapy (iCBT) with waitlist control (WLC). GAD: generalized anxiety disorder; iCBGT: clinician-guided group iCBT; PD: panic disorder; SAD: social anxiety disorder.



Group by Trial	Study name	Hedges *g*	Lower limit	Upper limit	*P* Value
	Bell, 2012 [82]	0.008	-0.465	0.482	0.972
	Kok, 2014 [50]	0.206	-0.064	0.476	0.134
	Mathiasen, 2016 [85]	0.369	-0.110	0.847	0.131
	Norgren, 2014 [7]	0.573	0.176	0.970	0.005
Clinical recruitment		0.281	0.095	0.466	0.003
	Andersson, 2006 [58]	0.579	0.085	1.073	0.022
	Andersson, 2012a [56]	0.495	-0.039	1.029	0.069
	Andersson, 2012b [59]	0.607	0.328	0.887	0.000
	Berger, 2009 [60]	0.880	0.309	1.451	0.003
	Berger, 2014 standardized [61]	0.906	0.471	1.341	0.000
	Berger, 2014 tailored [61]	0.857	0.424	1.290	0.000
	Berger, 2017 [31]	0.396	0.062	0.730	0.020
	Botella, 2010 [62]	0.531	0.088	0.975	0.019
	Carlbring, 2001 [63]	1.186	0.533	1.839	0.000
	Carlbring, 2006 [65]	1.368	0.812	1.924	0.000
	Carlbring, 2007 [66]	0.734	0.200	1.268	0.007
	Carlbring, 2011 [67]	0.376	-0.154	0.907	0.165
	Christensen, 2014 [68]	1.014	-0.073	2.101	0.067
	Furmark, 2009 [69]	0.767	0.317	1.217	0.001
	Gallego, 2011 [52]	0.384	-0.413	1.180	0.345
	Johnston, 2011 clinician [28]	0.814	0.392	1.236	0.000
	Johnston, 2011 coach [28]	1.215	0.776	1.654	0.000
	Klein, 2006 [71]	0.383	-0.254	1.020	0.239
	Oromendia, 2016 non-scheduled [27]	0.791	0.234	1.348	0.005
	Oromendia, 2016 scheduled [27]	1.552	0.927	2.177	0.000
	Paxling, 2011 [72]	0.844	0.414	1.274	0.000
	Richards, 2006 iCBT [73]	-0.317	-1.152	0.518	0.457
	Richards, 2006 iCBT&stressmgmt [73]	0.634	-0.232	1.500	0.151
	Robinson, 2010 clinician [74]	1.199	0.774	1.625	0.000
	Robinson, 2010 technician [74]	1.265	0.834	1.696	0.000
	Schulz, 2016 iCBGT [75]	0.774	0.320	1.228	0.001
	Schulz, 2016 iCBT [75]	1.011	0.547	1.475	0.000
	Titov, 2008a [76]	1.035	0.618	1.451	0.000
	Titov, 2008b [77]	1.085	0.641	1.529	0.000
	Titov, 2008c guided [78]	1.053	0.547	1.559	0.000
	Titov, 2008c self-help [78]	0.406	-0.077	0.889	0.099
	Titov, 2009 [79]	1.217	0.589	1.845	0.000
	Titov, 2010 GAD [30]	0.095	-0.563	0.753	0.777
	Titov, 2010 PD [30]	1.254	0.349	2.159	0.007
	Titov, 2010 SAD [30]	0.436	-0.362	1.235	0.284
	Titov, 2011 [29]	0.267	-0.376	0.910	0.416
	Wims, 2010 [80]	0.587	0.070	1.103	0.026
Open recruitment		0.790	0.708	0.872	0.000
Overall		0.707	0.632	0.782	0.000

Figure 3. Forest plot of effects on anxiety symptoms of open recruitment trials and clinical service recruitment trials comparing internet-delivered cognitive behavioral therapy (iCBT) with face-to-face cognitive behavioral therapy (CBT). PD: panic disorder; SAD: social anxiety disorder.

Discussion

Principal Findings

This meta-analysis showed that iCBT is more effective than WLC in reducing anxiety symptoms at posttreatment. We found no indication for differences in effect sizes between iCBT and face-to-face CBT. These outcomes confirm the results of previous meta-analyses, which found moderate to large effect sizes for WLC comparator studies and small and nonsignificant effect sizes when comparing iCBT versus face-to-face CBT [1-4].

Our main research question was whether OR trials produce effects for anxiety symptoms similar to those of CSR trials. For studies with WLC comparators, recruitment method was significantly associated with anxiety treatment outcomes (Q=10.09; *P*=.001), indicating that effect sizes are higher in OR trials than in CSR trials.

We explored whether differences between characteristics of samples in OR trials and CSR trials might explain the gap between effects we found in favor of OR trials. Multivariate metaregression analysis revealed that the association between recruitment method and effect size may be partly explained by greater treatment adherence and the exclusion of severely depressed patients in OR trials.

For studies with face-to-face CBT comparators, we observed no difference in anxiety outcomes between OR trials and CSR trials (Q=0.82; *P*=.37). Possibly, the number of studies with face-to-face CBT comparators was too low, making these analyses underpowered to detect differences. Another explanation could be that studies with face-to-face CBT comparators resembled each other more on other criteria for clinical representativeness than studies with WLC comparators, as these studies are conducted in routine care. For example, in studies with face-to-face comparators, treatments are generally delivered by skilled clinicians and in clinically representative settings. These study characteristics are more varied in studies with WLC comparators, where treatments can also be delivered by researchers or graduate students and in a research setting such as a university laboratory [108].

Regarding depressive symptoms and quality of life, we observed no differences between OR trials and CSR trials for either comparator group.

The difference in results we found between OR and CSR trials for studies with WLC comparators is in line with a previous meta-analysis on effectiveness of face-to-face CBT for anxiety disorders by Stewart and Chambless [9]. The small but significant effect size (*d*=–0.08; *P*<.05) they found indicated smaller improvements in more clinically representative patient studies than in less clinically representative studies.

Our findings are partly in keeping with Andersson and Hedman’s review on the effectiveness of iCBT for anxiety [21]. Results of that review suggested that effectiveness studies obtain similar effects to efficacy trials. Considering they only included studies comparing iCBT with face-to-face CBT, that conclusion corresponds to our results for iCBT compared with face-to-face CBT. However, it needs to be noted that Andersson and Hedman based their distinction between efficacy and effectiveness on the setting in which iCBT was delivered and not on recruitment strategy.

Strengths and Limitations

A major strength of this study is that it is, to our knowledge, the first meta-analysis of iCBT for anxiety disorders comparing treatment outcomes between OR and CSR for both WLC and face-to-face CBT comparators. Furthermore, the studies comparing iCBT versus face-to-face CBT were head-to-head comparisons, generating direct evidence.

Some limitations in this study warrant caution in interpretation. First, the number of trials was relatively low for studies with face-to-face CBT comparators. Hence, finding no difference may have been caused by underpowered analyses.

Second, clinical representativeness of studies is often rated based on a multitude of criteria, besides recruitment type, such as setting of treatment delivery, experience of therapists, and flexibility in treatment manuals [9,109]. This means that the differences we found may have been caused by predictors not assessed in this study. In future research, considering multiple criteria of efficacy and effectiveness would be helpful to more thoroughly determine clinical representativeness of the studies and the association between clinical representativeness and treatment outcomes.

Third, in 3 included trials a mixed recruitment strategy was applied. This contamination may have led to some bias. Any such bias will have decreased the difference between OR and CSR trials found in the meta-analysis. We decided to include these trials because they reported clearly on their recruitment method and also the portion of participants recruited through an OR method versus a CSR method.

Fourth, when interpreting the subgroup analyses and metaregression analyses, it is important to bear in mind that the results were only observational. Direct comparisons are required to verify the findings presented here.

Conclusions

This meta-analysis indicated that the effects of iCBT for anxiety disorders compared with WLC in CSR trials were smaller than effects found in OR trials. Hence, for patients with anxiety disorders in routine care, the impact of iCBT may not be as positive as for self-referred study participants recruited from the community. The difference between OR and CSR might be partly caused by a greater treatment adherence of self-referred patients and stricter exclusion criteria for severe depressive symptoms in studies with an OR method. A future challenge is to build a more robust body of evidence supporting the effectiveness of iCBT for anxiety disorders in routine care populations.

Conflicts of Interest

NT is Executive Director of MindSpot, which is funded by the Australian Government to deliver iCBT to adults with anxiety and depression across Australia.

Multimedia Appendix 1

Search string.

[PDF File (Adobe PDF File), 95KB - jmir_v21i4e11706_app1.pdf]

Multimedia Appendix 2

Order of rating scales.

[PDF File (Adobe PDF File), 71KB - jmir_v21i4e11706_app2.pdf]

Multimedia Appendix 3

Characteristics of included studies.

[PDF File (Adobe PDF File), 151KB - jmir_v21i4e11706_app3.pdf]

Multimedia Appendix 4

Risk-of-bias assessment.

[PDF File (Adobe PDF File), 327KB - jmir_v21i4e11706_app4.pdf]

Multimedia Appendix 5

Secondary outcomes.

[PDF File (Adobe PDF File), 16KB - jmir_v21i4e11706_app5.pdf]

Multimedia Appendix 6

Funnel plot publication bias.

[PDF File (Adobe PDF File), 97KB - jmir_v21i4e11706_app6.pdf]

Multimedia Appendix 7

Metaregression and subgroup analyses.

[PDF File (Adobe PDF File), 14KB - jmir_v21i4e11706_app7.pdf]

References

1. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565. [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
2. Adelman CB, Panza KE, Bartley CA, Bontempo A, Bloch MH. A meta-analysis of computerized cognitive-behavioral therapy for the treatment of DSM-5 anxiety disorders. J Clin Psychiatry 2014 Jul;75(7):e695-e704. [doi: 10.4088/JCP.13r08894] [Medline: 25093485]
3. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
4. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
5. Andersson G, Rozental A, Shafran R, Carlbring P. Long-term effects of internet-supported cognitive behaviour therapy. Expert Rev Neurother 2018 Jan;18(1):21-28. [doi: 10.1080/14737175.2018.1400381] [Medline: 29094622]
6. Hedman E, El Alaoui S, Lindefors N, Andersson E, Rück C, Ghaderi A, et al. Clinical effectiveness and cost-effectiveness of internet- vs. group-based cognitive behavior therapy for social anxiety disorder: 4-year follow-up of a randomized trial. Behav Res Ther 2014 Aug;59:20-29. [doi: 10.1016/j.brat.2014.05.010] [Medline: 24949908]
7. Nordgren LB, Hedman E, Etienne J, Bodin J, Kadowaki A, Eriksson S, et al. Effectiveness and cost-effectiveness of individually tailored internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. Behav Res Ther 2014 Aug;59:1-11 [FREE Full text] [doi: 10.1016/j.brat.2014.05.007] [Medline: 24933451]
8. Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. J Consult Clin Psychol 2001 Dec;69(6):875-899. [Medline: 11777114]
9. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009 Aug;77(4):595-606. [doi: 10.1037/a0016032] [Medline: 19634954]
10. Streiner DL. The 2 “Es” of research: efficacy and effectiveness trials. Can J Psychiatry 2002 Aug;47(6):552-556. [doi: 10.1177/070674370204700607] [Medline: 12211883]
11. Persons JB, Silberschatz G. Are results of randomized controlled trials useful to psychotherapists? J Consult Clin Psychol 1998 Feb;66(1):126-135. [Medline: 9489266]
12. Stirman SW, DeRubeis RJ, Crits-Christoph P, Brody PE. Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. J Consult Clin Psychol 2003 Dec;71(6):963-972. [doi: 10.1037/0022-006X.71.6.963] [Medline: 14622071]

13. Titov N, Andrews G, Kemp A, Robinson E. Characteristics of adults with anxiety or depression treated at an internet clinic: comparison with a national survey and an outpatient clinic. PLoS One 2010;5(5):e10885 [FREE Full text] [doi: 10.1371/journal.pone.0010885] [Medline: 20526371]
14. Wilks CR, Zieve GG, Lessing HK. Are trials of computerized therapy generalizable? A multidimensional meta-analysis. Telemed J E Health 2016 May;22(5):450-457. [doi: 10.1089/tmj.2015.0129] [Medline: 26461235]
15. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Terides MD, et al. Examining self-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety and depression: Two feasibility open trials. Internet Interventions 2015 Mar;2(1):17-23. [doi: 10.1016/j.invent.2014.11.002]
16. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. The first 30 months of the MindSpot Clinic: evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
17. Hadjistavropoulos HD, Pugh NE, Hesser H, Andersson G. Predicting response to therapist-assisted internet-delivered cognitive behavior therapy for depression or anxiety within an open dissemination trial. Behav Ther 2016 Mar;47(2):155-165. [doi: 10.1016/j.beth.2015.10.006] [Medline: 26956649]
18. Ruwaard J, Lange A, Schrieken B, Dolan CV, Emmelkamp P. The effectiveness of online cognitive behavioral treatment in routine clinical practice. PLoS One 2012;7(7):e40089 [FREE Full text] [doi: 10.1371/journal.pone.0040089] [Medline: 22792217]
19. Williams AD, O'Moore K, Mason E, Andrews G. The effectiveness of internet cognitive behaviour therapy (iCBT) for social anxiety disorder across two routine practice pathways. Internet Interv 2014 Oct;1(4):225-229. [doi: 10.1016/j.invent.2014.11.001]
20. Mathiasen K, Riper H, Andersen TE, Roessler KK. Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: observational study. J Med Internet Res 2018 Nov 28;20(11):e10927 [FREE Full text] [doi: 10.2196/10927] [Medline: 30487118]
21. Andersson G, Hedman E. Effectiveness of guided Internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013;23(3):140-148. [doi: 10.1159/000354779]
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
23. Riper H, Ballegooijen WV, Kooistra L. Preventie & eMental-health. Onderzoek dat leidt, technologie die verleidt, preventie die bereikt en beklijft. Kennissynthese 2013 in opdracht van ZonMw. Amsterdam, the Netherlands: ZonMw; 2013.
24. Mahoney AEJ, Mackenzie A, Williams AD, Smith J, Andrews G. Internet cognitive behavioural treatment for obsessive compulsive disorder: a randomised controlled trial. Behav Res Ther 2014 Dec;63:99-106. [doi: 10.1016/j.brat.2014.09.012] [Medline: 25461784]
25. Engel CC, Litz B, Magruder KM, Harper E, Gore K, Stein N, et al. Delivery of self training and education for stressful situations (DESTRESS-PC): a randomized trial of nurse assisted online self-management for PTSD in primary care. Gen Hosp Psychiatry 2015;37(4):323-328 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.04.007] [Medline: 25929985]
26. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, NY: Guilford Press; 1979.
27. Oromendia P, Orrego J, Bonillo A, Molinuevo B. Internet-based self-help treatment for panic disorder: a randomized controlled trial comparing mandatory versus optional complementary psychological support. Cogn Behav Ther 2016 Jun;45(4):270-286. [doi: 10.1080/16506073.2016.1163615] [Medline: 27007256]
28. Johnston L, Titov N, Andrews G, Spence J, Dear BF. A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]
29. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
30. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic Internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
31. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
32. Heading K, Kirkby KC, Martin F, Daniels BA, Gilroy LJ, Menzies RG. Controlled comparison of single-session treatments for spider phobia: live graded exposure alone versus computer-aided vicarious exposure. Behav Change 2012 Feb 22;18(02):103-113. [doi: 10.1375/bech.18.2.103]
33. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
34. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized

controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]

35. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [Medline: 3204199]
36. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia; 1995.
37. Spielberger CD, Gorsuch RL, Lushene RE. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1970:1-23.
38. Taylor S, Cox B. Anxiety sensitivity: multiple dimensions and hierarchic structure. Behav Res Ther 1998 Jan;36(1):37-51. [Medline: 9613015]
39. Liebowitz MR. Social phobia. Mod Probl Pharmacopsychiatry 1987;22:141-173. [Medline: 2885745]
40. Heimberg R, Horner K, Juster H, Safren S, Brown E, Schneier F, et al. Psychometric properties of the Liebowitz Social Anxiety Scale. Psychol Med 1999 Jan;29(1):199-212. [Medline: 10077308]
41. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [Medline: 9670605]
42. Furmark T, Tillfors M, Everz P, Marteinsdottir I, Gefvert O, Fredrikson M. Social phobia in the general population: prevalence and sociodemographic profile. Soc Psychiatry Psychiatr Epidemiol 1999 Aug;34(8):416-424. [Medline: 10501711]
43. Leary M. A brief version of the Fear of Negative Evaluation Scale. Personal Soc Psychol Bull 1983;9(3):371-375. [Medline: 803973233]
44. 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]
45. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [Medline: 2076086]
46. Shear MK, Brown T, Barlow D, Money R, Sholomskas D, Woods S, et al. Multicenter collaborative panic disorder severity scale. Am J Psychiatry 1997 Nov;154(11):1571-1575. [doi: 10.1176/ajp.154.11.1571] [Medline: 9356566]
47. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. J Consult Clin Psychol 1984 Dec;52(6):1090-1097. [Medline: 6520279]
48. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The Panic and Agoraphobia Scale. Int Clin Psychopharmacol 1995 Jun;10(2):73-81. [Medline: 7673659]
49. Van Zuuren FJ. The fear questionnaire. Some data on validity, reliability and layout. Br J Psychiatry 1988 Nov;153:659-662. [Medline: 3255455]
50. Kok R, van Straten A, Beekman A, Cuijpers P. Short-term effectiveness of web-based guided self-help for phobic outpatients: randomized controlled trial. J Med Internet Res 2014 Sep 29;16(9):e226 [FREE Full text] [doi: 10.2196/jmir.3429] [Medline: 25266929]
51. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
52. Gallego M, Gerardus EP, van der Kooij M, Mees H. The effects of a Dutch version of an internet-based treatment program for fear of public speaking: a controlled study. Int J Clin Heal Psychol 2011;11(3):459-472 [FREE Full text]
53. 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;34(1):9-17. [Medline: 14971623]
54. Nordgreen T, Haug T, Öst L, Andersson G, Carlbring P, Kvale G, et al. Stepped care versus direct face-to-face cognitive behavior therapy for social anxiety disorder and panic disorder: a randomized effectiveness trial. Behav Ther 2016 Mar;47(2):166-183. [doi: 10.1016/j.beth.2015.10.004] [Medline: 26956650]
55. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost L. Internet-based self-help versus one-session exposure in the treatment of spider phobia: a randomized controlled trial. Cogn Behav Ther 2009;38(2):114-120. [doi: 10.1080/16506070902931326] [Medline: 20183690]
56. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom 2012;81(6):344-355. [doi: 10.1159/000339371] [Medline: 22964540]
57. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost LG. Internet-based exposure treatment versus one-session exposure treatment of snake phobia: a randomized controlled trial. Cogn Behav Ther 2013;42(4):284-291. [doi: 10.1080/16506073.2013.844202] [Medline: 24245707]
58. Andersson G, Carlbring P, Holmström A, Sparthan E, Furmark T, Nilsson-Ihrfelt E, et al. Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: a randomized controlled trial. J Consult Clin Psychol 2006 Aug;74(4):677-686. [doi: 10.1037/0022-006X.74.4.677] [Medline: 16881775]

59. Andersson G, Carlbring P, Furmark T, S. O. F. I. E. Research Group. Therapist experience and knowledge acquisition in internet-delivered CBT for social anxiety disorder: a randomized controlled trial. PLoS One 2012;7(5):e37411 [FREE Full text] [doi: 10.1371/journal.pone.0037411] [Medline: 22649526]
60. Berger T, Hohl E, Caspar F. Internet-based treatment for social phobia: a randomized controlled trial. J Clin Psychol 2009 Oct;65(10):1021-1035. [doi: 10.1002/jclp.20603] [Medline: 19437505]
61. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
62. Botella C, Gallego MJ, Garcia-Palacios A, Guillen V, Baños RM, Quero S, et al. An Internet-based self-help treatment for fear of public speaking: a controlled trial. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):407-421. [doi: 10.1089/cyber.2009.0224] [Medline: 20712499]
63. Carlbring P, Westling BE, Ljungstrand P, Ekselius L, Andersson G. Treatment of panic disorder via the internet: a randomized trial of a self-help program. Behav Ther 2001;32(4):751-764. [doi: 10.1016/S0005-7894(01)80019-8]
64. Carlbring P, Nilsson-Ihrfelt E, Waara J, Kollenstam C, Buhrman M, Kaldo V, et al. Treatment of panic disorder: live therapy vs. self-help via the Internet. Behav Res Ther 2005 Oct;43(10):1321-1333. [doi: 10.1016/j.brat.2004.10.002] [Medline: 16086983]
65. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/appi.ajp.163.12.2119] [Medline: 17151163]
66. Carlbring P, Gunnarsdóttir M, Hedensjö L, Andersson G, Ekselius L, Furmark T. Treatment of social phobia: randomised trial of internet-delivered cognitive-behavioural therapy with telephone support. Br J Psychiatry 2007 Feb;190:123-128. [doi: 10.1192/bjp.bp.105.020107] [Medline: 17267928]
67. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, Internet-based treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002] [Medline: 21047620]
68. Christensen H, Mackinnon A, Batterham P, O'Dea B, Guastella A, Griffiths K, et al. The effectiveness of an online e-health application compared to attention placebo or Sertraline in the treatment of Generalised Anxiety Disorder. Internet Interventions 2014 Oct;1(4):169-174. [doi: 10.1016/j.invent.2014.08.002]
69. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
70. Kiropoulos LA, Klein B, Austin DW, Gilson K, Pier C, Mitchell J, et al. Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? J Anxiety Disord 2008 Dec;22(8):1273-1284. [doi: 10.1016/j.janxdis.2008.01.008] [Medline: 18289829]
71. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]
72. Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, et al. Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: a randomized controlled trial. Cogn Behav Ther 2011;40(3):159-173. [doi: 10.1080/16506073.2011.576699] [Medline: 21770848]
73. Richards JC, Klein B, Austin DW. Internet cognitive behavioural therapy for panic disorder: does the inclusion of stress management information improve end-state functioning? Clin Psychol 2006 Mar;10(1):2-15. [doi: 10.1080/13284200500378795]
74. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
75. Schulz A, Stolz T, Vincent A, Krieger T, Andersson G, Berger T. A sorrow shared is a sorrow halved? A three-arm randomized controlled trial comparing internet-based clinician-guided individual versus group treatment for social anxiety disorder. Behav Res Ther 2016 Sep;84:14-26. [doi: 10.1016/j.brat.2016.07.001] [Medline: 27423374]
76. Titov N, Andrews G, Schwencke G, Drobny J, Einstein D. Shyness 1: distance treatment of social phobia over the Internet. Aust N Z J Psychiatry 2008 Jul;42(7):585-594. [doi: 10.1080/00048670802119762] [Medline: 18612862]
77. Titov N, Andrews G, Schwencke G. Shyness 2: treating social phobia online: replication and extension. Aust N Z J Psychiatry 2008 Jul;42(7):595-605. [doi: 10.1080/00048670802119820] [Medline: 18612863]
78. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
79. Titov N, Andrews G, Robinson E, Schwencke G, Johnston L, Solley K, et al. Clinician-Assisted Internet-Based Treatment is Effective for Generalized Anxiety Disorder: Randomized Controlled Trial. Aust N Z J Psychiatry 2009 Jan;43(10):905-912. [doi: 10.1080/00048670903179269]

80. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: a randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]
81. Andrews G, Davies M, Titov N. Effectiveness randomized controlled trial of face to face versus Internet cognitive behaviour therapy for social phobia. Aust N Z J Psychiatry 2011 Apr;45(4):337-340. [doi: 10.3109/00048674.2010.538840] [Medline: 21323490]
82. Bell C, Colhoun H, Carter F, Frampton C. Effectiveness of computerised cognitive behaviour therapy for anxiety disorders in secondary care. Aust N Z J Psychiatry 2012 Jul;46(7):630-640. [doi: 10.1177/0004867412437345] [Medline: 22327097]
83. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. [An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences]. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]
84. Hedman E, Andersson G, Ljótsson B, Andersson E, Rück C, Mörtberg E, et al. Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: a randomized controlled non-inferiority trial. PLoS One 2011;6(3):e18001 [FREE Full text] [doi: 10.1371/journal.pone.0018001] [Medline: 21483704]
85. Mathiasen K, Riper H, Ehlers L, Valentin J, Rosenberg N. Internet-based CBT for social phobia and panic disorder in a specialised anxiety clinic in routine care: results of a pilot randomised controlled trial. Internet Interv 2016 May;4:92-98 [FREE Full text] [doi: 10.1016/j.invent.2016.03.001] [Medline: 30135794]
86. Norberg M, Diefenbach G, Tolin D. Quality of life and anxiety and depressive disorder comorbidity. J Anxiety Disord 2008 Dec;22(8):1516-1522. [doi: 10.1016/j.janxdis.2008.03.005] [Medline: 18424062]
87. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
88. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [Medline: 8140903]
89. 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] [Medline: 11556941]
90. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
91. Frisch M, Cornell J, Villanueva M, Retzlaff P. Clinical validation of the Quality of Life Inventory. A measure of life satisfaction for use in treatment planning and outcome assessment. Psychological Assessment 1992;4(1):92-101. [doi: 10.1037/1040-3590.4.1.92]
92. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May;28(3):551-558. [Medline: 9626712]
93. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
94. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]
95. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Intervention. Chichester, UK: Wiley-Blackwell Publishing Ltd; 2008.
96. Hedges LV, Olkin I. Statistical Methods For Meta-analysis. Orlando, FL: Academic Press; 1995.
97. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum; 1988.
98. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
99. Borenstein M, Higgins J, Hedges LV, Rothstein HR. Basics of meta-analysis: I is not an absolute measure of heterogeneity. Res Synth Methods 2017 Mar;8(1):5-18. [doi: 10.1002/jrsm.1230] [Medline: 28058794]
100. Orsini N, Bottai M, Higgins J, Buchan I. Heterogi: Stata Module to Quantify Heterogeneity in a Meta-Analysis. Chestnut Hill, MA: Department of Economics, Boston College; 2006.
101. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to Meta-Analysis. Oxford, UK: Wiley; 2009.
102. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [Medline: 9310563]
103. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [Medline: 10877304]
104. Allison P. Logistic Regression Using SAS: Theory And Application. 1st edition. Cary, NC: SAS Institute; 1999.
105. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 2006 Jun 1;59(11):990-996. [doi: 10.1016/j.biopsych.2005.09.014] [Medline: 16368078]
106. Citrome L, Ketter T. When does a difference make a difference? Interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. Int J Clin Pract 2013 May;67(5):407-411. [doi: 10.1111/ijcp.12142] [Medline: 23574101]
107. Klein B, Richards JC. A brief internet-based treatment for panic disorder. Behav Cogn Psychother 2001 Mar 6;29(01):113-117. [doi: 10.1017/S1352465801001138]

108. Weisz J, Donenberg G, Han S, Weiss B. Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. J Consult Clin Psychol 1995 Oct;63(5):688-701. [Medline: 7593861]
109. Shadish WR, Matt GE, Navarro AM, Siegle G, Crits-Christoph P, Hazelrigg MD, et al. Evidence that therapy works in clinically representative conditions. J Consult Clin Psychol 1997 Jun;65(3):355-365. [Medline: 9170759]

Abbreviations

BAI: Beck Anxiety Inventory
CBT: cognitive behavioral therapy
CSR: clinical service recruitment
DSM: Diagnostic and Statistical Manual of Mental Disorders
iCBT: internet-delivered cognitive behavioral therapy
NNT: number needed to treat
OR: open recruitment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WLC: waitlist control

Edited by G Eysenbach; submitted 26.07.18; peer-reviewed by S Bennett, G Zieve; comments to author 16.08.18; revised version received 31.01.19; accepted 24.03.19; published 17.04.19.

Please cite as:
Romijn G, Batelaan N, Kok R, Koning J, van Balkom A, Titov N, Riper H
Internet-Delivered Cognitive Behavioral Therapy for Anxiety Disorders in Open Community Versus Clinical Service Recruitment: Meta-Analysis
J Med Internet Res 2019;21(4):e11706
URL: http://www.jmir.org/2019/4/e11706/
doi:10.2196/11706
PMID:30994462

Original Paper

Using Machine Learning to Derive Just-In-Time and Personalized Predictors of Stress: Observational Study Bridging the Gap Between Nomothetic and Ideographic Approaches

Alan Rozet[1], BA; Ian M Kronish[1], MPH, MD; Joseph E Schwartz[1], MS, PhD; Karina W Davidson[2], MASc, PhD

[1]Center for Behavioral Cardiovascular Health, Columbia University Irving Medical Center, New York, NY, United States

[2]Feinstein Institute for Medical Research, Northwell Health, New York, NY, United States

Corresponding Author:
Alan Rozet, BA
Center for Behavioral Cardiovascular Health
Columbia University Irving Medical Center
Presbyterian Hospital Building, 9th Floor
622 W 168th Street
New York, NY, 10032
United States
Phone: 1 212 342 4493
Email: ar3793@cumc.columbia.edu

Abstract

Background: Investigations into person-specific predictors of stress have typically taken either a population-level nomothetic approach or an individualized ideographic approach. Nomothetic approaches can quickly identify predictors but can be hindered by the heterogeneity of these predictors across individuals and time. Ideographic approaches may result in more predictive models at the individual level but require a longer period of data collection to identify robust predictors.

Objective: Our objectives were to compare predictors of stress identified through nomothetic and ideographic models and to assess whether sequentially combining nomothetic and ideographic models could yield more accurate and actionable predictions of stress than relying on either model. At the same time, we sought to maintain the interpretability necessary to retrieve individual predictors of stress despite using nomothetic models.

Methods: Data collected in a 1-year observational study of 79 participants performing low levels of exercise were used. Physical activity was continuously and objectively monitored by actigraphy. Perceived stress was recorded by participants via daily ecological momentary assessments on a mobile app. Environmental variables including daylight time, temperature, and precipitation were retrieved from the public archives. Using these environmental, actigraphy, and mobile assessment data, we built machine learning models to predict individual stress ratings using linear, decision tree, and neural network techniques employing nomothetic and ideographic approaches. The accuracy of the approaches for predicting individual stress ratings was compared based on classification errors.

Results: Across the group of patients, an individual's recent history of stress ratings was most heavily weighted in predicting a future stress rating in the nomothetic recurrent neural network model, whereas environmental factors such as temperature and daylight, as well as duration and frequency of bouts of exercise, were more heavily weighted in the ideographic models. The nomothetic recurrent neural network model was the highest performing nomothetic model and yielded 72% accuracy for an 80%/20% train/test split. Using the same 80/20 split, the ideographic models yielded 75% accuracy. However, restricting ideographic models to participants with more than 50 valid days in the training set, with the same 80/20 split, yielded 85% accuracy.

Conclusions: We conclude that for some applications, nomothetic models may be useful for yielding higher initial performance while still surfacing personalized predictors of stress, before switching to ideographic models upon sufficient data collection.

(J Med Internet Res 2019;21(4):e12910) doi:10.2196/12910

KEYWORDS

ecological momentary assessment; machine learning; stress-behavior pathway; personal informatics; self-quantification; exercise; weather; just-in-time interventions

Introduction

Deeper knowledge of the day-to-day effects of both weather and physical activity on stress can be valuable for creating personalized stress-reduction interventions on a just-in-time basis. Previous investigations have often focused on a nomothetic approach, pooling data to identify influential features across individuals [1,2]. However, this approach typically has a drawback: Insight into any particular individual is limited due to heterogeneous effects of factors on individual-level stress and may not be generalized due to biological variability or overfitting [3]. For example, hot weather may reduce stress for some participants, but increase stress for others. To remedy this, some have taken an ideographic approach, developing personalized stress-prediction models for each person [4-6].

Although this ideographic strategy resolves the issue of achieving per-individual insights, it does so by discarding potentially useful data from other individuals. The ideographic approach also requires the acquisition of at least some data about a given individual before making predictions for that individual. Identifying predictors of potentially low-frequency events such as occurrences of high stress may require substantial data collection before the ideographic model becomes sufficiently robust to confidently identify predictors. Thus, prior to obtaining reliable predictors of stress, a given individual may have to complete a long period of self-tracking, which may not be acceptable to some individuals.

Here, we first compared individual-level predictors of stress identified through nomothetic models to those identified in ideographic models. We next explored the accuracy of a model-switching paradigm that begins with a nomothetic model and progressively changes to an ideographic model for data for the individual accumulate. We hypothesized that beginning with a nomothetic model would maximize accuracy during the early phase of data collection (referred to as a “warm start”) and subsequently, switching to ideographic modeling for higher personalization and performance would be the most effective approach to maximizing accuracy throughout data collection.

Methods

Overview

This paper used the data collected in an observational study of 79 participants who were followed for up to 1 year, starting between January and July 2014; the study continuously and objectively monitored the physical activity of the participants by actigraphy and the perceived stress rating through ecological momentary assessment (EMA) reports on a mobile app [6]. Participants were healthy individuals, aged 18 years or older, who responded to fliers posted throughout the buildings of Columbia University Medical Center (New York City, NY) and who, on phone screening, reported only intermittent engagement in exercise and having access to a personal computer and iOS or Android smartphone. Individuals with significant medical comorbidities, occupational work demands requiring rigorous physical activity, or inability to read and speak English were excluded. During a baseline interview, demographic characteristics including age, sex, race, ethnicity, education, partner status, and living situation were collected.

Measures

Stress was measured using an end-of-day text message survey on the participant’s own iPhone or Android phone, with the question “Overall, how stressful was your day?” Each evening, the participant was also asked, “Overall, how stressful do you think tomorrow will be?” Each morning, the participant responded to the questions “How stressful do you expect today to be?” and “How likely are you to exercise today?” All responses were rated on a scale from 0 (not at all) to 10 (extremely). All surveys were administered using Qualtrics software (Qualtrics, Seattle, WA). Two participants’ data were excluded for almost no variance in the self-reported stress ratings, leading to a total of 77 participants for the analysis.

Physical activity was measured using a wrist-worn Fitbit (Fitbit, Inc, San Francisco, CA) to track daily physical activity, including the steps taken, calories burned, and intensity of physical activity for each minute of the day. Participants were instructed to sync and charge the device every 5 to 7 days. In this analysis, a bout of “exercise” was defined as any consecutive 30-minute period within which 24 or more minutes of moderate- or vigorous-intensity activity was performed. We followed the recommendations of Ward et al [7] regarding best practices for the use of accelerometer data in research on physical activity. Specifically, physical activity guidelines recommend exercising for at least 30 minutes a day while accommodating interruptions. Further, when analyzing accelerometer data, the conventional approach is to quantify exercise in bouts of 10 minutes with allowances for 2 minutes of interruption (ie, total exercise for 8 of 10 min). Extrapolating the definition based on 10 to 30 minutes of activity, this yielded 24 of 30 minutes of activity. Software was written to determine, for each day, whether there was any 30-minute period within which at least 24 minutes of moderate or vigorous activity was performed; this was our objectively assessed measure of a 30-minute period of exercise. Days in which the Fitbit device was worn for fewer than 10 hours were excluded from all analyses.

External and environmental variables, including temperature (high, low, average, and range), hours of daylight, precipitation, and day of the week, were retrieved from the meteorological station in Central Park (New York City, NY); these data are made publicly available by the National Oceanic and Atmospheric Administration’s National Center for Environmental Information.

Statistical Analysis

We developed models for stress rating using neural network, decision tree, and linear approaches across all participants, as well as participant-specific decision trees. Each model predicted a person’s self-reported stress rating (range, 0-10), using the previous 3 days of weather, self-reported stress, and actigraphy data. Because the previous 3 days were used as inputs, periods missing a stress rating were excluded.

The continuous prediction for stress was then converted into a binary classification as either above or below a participant’s median self-reported stress value in the training set. We chose this approach, because if the stress rating was left as a continuous value, it was not clear at what rating an intervention should be initiated. Further, using the participant’s median as a reference slightly adjusts for participants who did not utilize the full range of the 11-point scale. In this way, the high-stress rating was individualized for each participant. The Keras Python library [8] was used to train neural network models, and the scikit-learn Python library [9] was used to train linear and decision tree-type models. Dropout, a regularization method for neural networks, was also used in the neural network models during training to decrease overfitting.

Model performance was compared based on regression (mean absolute error) and classification (area under the curve [AUC], F_1 score, accuracy) error in predicting stress self-report. Feature importances were also compared between models. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [10]. A model with an AUC closer to 1 is generally better. An F_1 score is the harmonic mean of a model’s precision and recall, with values between 0 and 1, in which values closer to 1 are better. Model selection was performed through exhaustive grid search of the corresponding hyperparameters for each model using 3-fold cross-validation and optimizing on the mean absolute error.

For the ideographic decision trees, the Gini importance, as implemented in scikit-learn, was used to derive the significance of each input variable for each participant. Layer-wise relevance propagation (LRP) [11] was used to interpret our neural network models. LRP propagates the relevance of each input variable back through the network from a specific prediction. In other words, for this dataset, a participant with 30 valid days in the dataset would yield 30 sets of LRP relevance scores, each set having one score for each input variable. These scores can be positive or negative in a similar fashion as linear coefficients, and the greater the magnitude of the score, the more that particular variable increased or decreased that particular prediction for stress rating. LRP was used in favor of other methods such as Deep Learning Important Features (DeepLIFT) and Integrated Gradients, because it does not require a baseline or reference value. However, LRP can be viewed as an approximation of DeepLIFT when bias terms are 0 and the reference values are set to 0 [12]. This, in turn, approximates Shapley values [13], which is another method of interpreting neural network output. As such, to leverage the visualizations for Shapley values built in the Shapley Additive Explanations Python library and to add an additional method of visual interpretation, bias terms for the neural net were locked to 0; this did not impact model performance.

The following variables were used for the ideographic models: the number of hours with ≥10 activities, total exercise duration in minutes during bouts of exercise with ≥24 minutes of activity out of 30 minutes, the number of exercise bouts with ≥24 minutes of activity out of 30 minutes, the binary presence of any exercise bout of ≥24 minutes of activity out of 30 minutes, the total number of exercise bouts, the binary presence of any exercise bout, total exercise duration, stress rating, minimum outdoor temperature, maximum outdoor temperature, average outdoor temperature, total daylight minutes, and total precipitation. Several person-level variables were included in the nomothetic models that were not included in the per-participant models, as they would have been static for a given participant and would not have contributed to performance. These included height, weight, age, and sex. The inclusion of these variables in the nomothetic models slightly improved their performance.

Additionally, all models were trained with varying training set sizes to test the hypothesis that ideographic models may be more reliable at large-enough training set sizes for each participant. Training set sizes ranged from 30% to 80% of valid days for each participant. For example, the first 30% of days in which a participant documented his or her stress was used to train a model predicting the subsequent 70% of days. The results for an 80% train/20% test split are highlighted here, with a total of 4050 training set samples and 678 testing set samples available to all nomothetic models.

All models were compared to a baseline model that simply predicted the median of the stress self-report values from a participant’s training set, for all test samples.

Results

Overview

The average age of the participants was 32 years (range, 20-58 years), with a height of 169 cm and weight of 75 kg. The study sample included 34 men and 43 women in the dataset. Table 1 depicts the basic descriptive statistics on the age, height, and weight of our set of participants.

The pattern of stress ratings differed significantly among participants. Figure 1 shows plots of stress by day for the 10 participants with the greatest number of valid responses in the dataset. Days without responses are shown without points plotted.

Table 1. Basic summary statistics for the participants (N=77).

Statistic	Mean (SD)	Min	25%[a]	50%[a]	75%[a]	Max
Age (years)	31.62 (9.42)	20	24	27	38	58
Height (cm)	168.71 (8.49)	149.4	162	168	176	187
Weight (kg)	75.68 (17.46)	43.3	62.5	74.7	85.5	138.6

[a] Quartile ranges.

Figure 1. Plots of stress ratings for the participants with the greatest amount of responses.

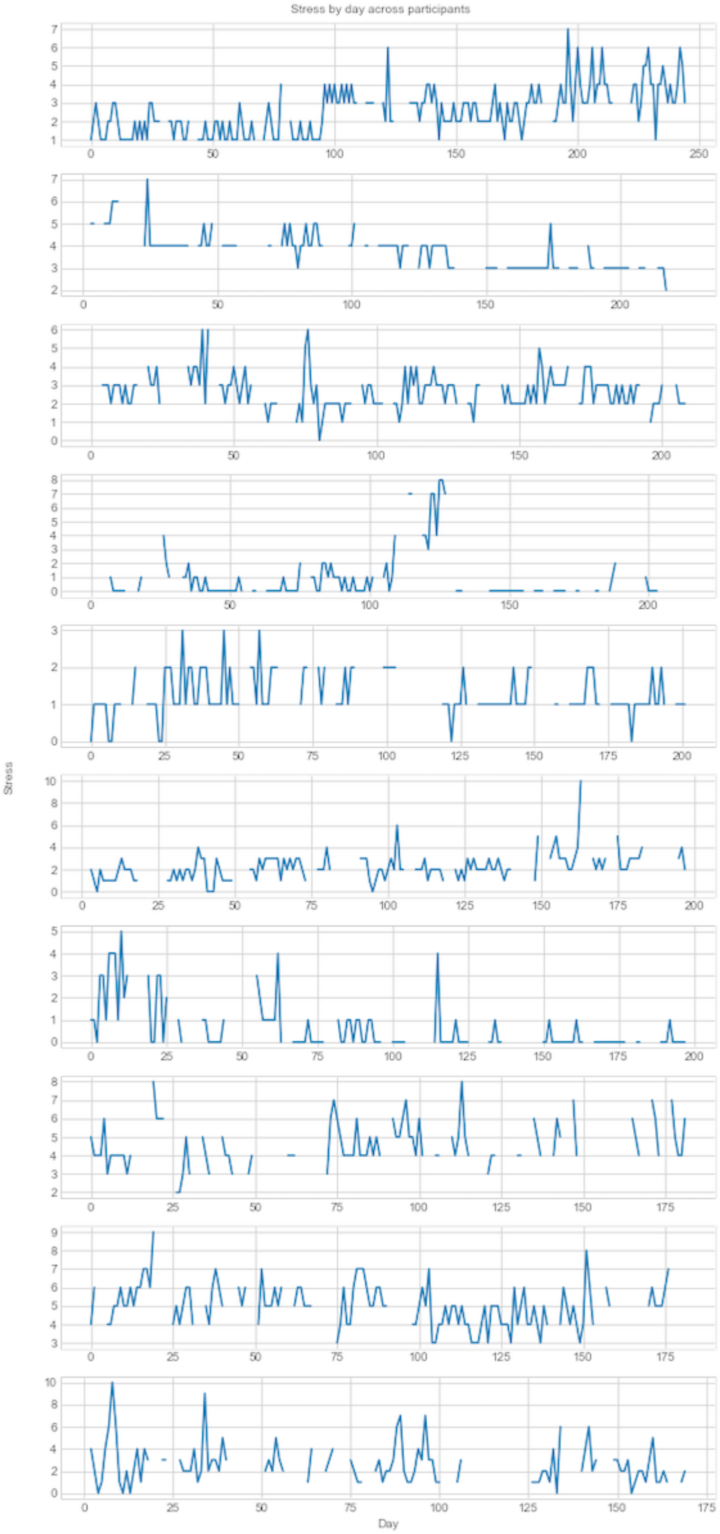


Table 2. Summary statistics and count of stress ratings, grouped by occurrence of missing stress ratings for the 3 days before the date of the predicted stress rating.

Response	Count	Mean stress (SD)	Min	25%[a]	50%[a]	75%[a]	Max
000	1728	3.60 (2.58)	0	2	3	6	10
001	1077	3.40 (2.57)	0	1	3	5	10
010	292	3.36 (2.63)	0	1	3	5	10
011	973	3.31 (2.51)	0	1	3	5	10
100	973	3.16 (2.54)	0	1	3	4	10
101	239	3.35 (2.63)	0	1	3	5	10
110	889	3.17 (2.48)	0	1	3	4	10
111	4060	2.93 (2.35)	0	1	3	4	10

[a] Quartile ranges.

Table 3. Percentage of stress ratings grouped by the occurrence of missing stress ratings for the 3 days before the date of the predicted stress rating. Each column corresponds to the value of the stress rating, and each cell represents what percentage of stress ratings had that value and response pattern.

Response	0	1	2	3	4	5	6	7	8	9	10
000	11.92	12.21	14.47	14.47	15.74	5.73	8.91	8.39	4.40	1.50	2.26
001	13.65	12.26	16.06	16.34	12.91	6.41	8.73	6.41	2.79	1.58	2.88
010	17.47	9.93	13.36	17.12	13.70	6.16	7.88	7.19	3.42	0.00	3.77
011	14.59	11.51	14.59	20.45	11.51	7.09	7.71	5.96	3.19	0.41	2.98
100	17.57	13.05	13.77	14.80	16.14	6.06	6.37	6.37	2.26	0.72	2.88
101	17.57	10.88	13.39	15.90	13.81	5.02	6.69	9.21	5.02	0.42	2.09
110	16.20	11.25	15.75	18.67	14.40	5.29	7.31	5.29	2.25	0.79	2.81
111	15.62	15.12	17.02	18.92	13.42	5.91	4.98	3.89	2.09	0.64	2.39

Table 2 shows the mean stress rating grouped by the occurrence of missing stress ratings for the 3 days before the date of the predicted stress rating. The response column describes the occurrence of stress ratings: 000 represents no stress ratings in any of the 3 days, 001 represents presence of only one stress rating recorded just before the predicted day, and 111 represents presence of all three stress ratings. Periods missing a stress rating were more often followed by a higher stress rating than periods that were not missing any stress ratings.

Table 3 shows the percentage of stress ratings grouped by the occurrence of missing stress ratings for the 3 days before the date of the predicted stress rating. The frequency of high-stress ratings is slightly higher for response patterns with missing stress ratings.

Accuracy of Nomothetic Versus Ideographic Models

With an 80% training set/20% testing set split, the nomothetic recurrent neural network model AUC was 74.20% and the F_1 score was 79.21%. In addition, the per-participant decision tree AUC was 0.67 and the F_1 score was 0.83. The full results can be found in Multimedia Appendix 1. Nomothetic models, on an average, outperformed ideographic models for shorter training set sizes. As the training set size increased (ie, allowing for longer duration of assessments), group-level model performance generally increased across all models, and the nomothetic and ideographic model performances converged. However, when separately applying models to individual participants, ideographic models generally outperformed the nomothetic ones once the days with valid data exceeded 50 days.

Table 4 lists performance metrics for each model, grouped by training set proportion and model used. Precision, recall, F_1 score, AUC, and accuracy are performance metrics. Training set size ranges from 30% to 80%. “Ideographic decision tree over 100 days” represents the performance metrics, only for participants who had at least 100 valid days, which is the same for the 50-day model.

Table 4. Performance metrics across all models, grouped by training set size.

Average number of days for training set size	Precision (%)	F_1 score (%)	Area under the curve	Accuracy (%)	Training set size (%)	Model
16.7215	74.22	71.95	53.39	60.69	30.00	Ideographic Decision Tree
16.7215	75.76	72.63	54.40	61.53	30.00	Ideographic Decision Tree Over 50 Days
16.7215	76.61	71.86	52.73	60.21	30.00	Ideographic Decision Tree Over 100 Days
16.7215	77.80	78.77	62.88	69.75	30.00	Random Forest
16.7215	78.83	78.72	64.25	70.10	30.00	Gradient Boosted Decision Tree
16.7215	79.90	77.93	65.32	69.69	30.00	Recurrent Neural Network
16.7215	81.09	72.69	64.71	65.18	30.00	Baseline
16.7215	81.20	81.03	68.21	73.36	30.00	Elastic Net
16.7215	82.41	77.39	68.00	70.01	30.00	Neural Network
21.7595	75.35	72.40	54.44	61.37	40.00	Ideographic Decision Tree Over 100 Days
21.7595	76.48	73.06	57.12	62.79	40.00	Ideographic Decision Tree
21.7595	76.96	72.96	56.32	62.37	40.00	Ideographic Decision Tree Over 50 Days
21.7595	77.87	77.09	62.08	67.96	40.00	Gradient Boosted Decision Tree
21.7595	78.15	79.23	63.16	70.25	40.00	Random Forest
21.7595	80.93	71.11	63.74	63.61	40.00	Baseline
21.7595	81.47	76.25	66.23	68.47	40.00	Recurrent Neural Network
21.7595	81.70	80.15	68.15	72.48	40.00	Elastic Net
21.7595	82.33	79.40	68.55	71.90	40.00	Neural Network
27.3291	75.76	75.36	52.41	63.47	50.00	Ideographic Decision Tree
27.3291	77.09	75.95	52.52	64.06	50.00	Ideographic Decision Tree Over 50 Days
27.3291	77.27	75.56	50.10	63.07	50.00	Ideographic Decision Tree Over 100 Days
27.3291	82.41	82.65	67.54	74.60	50.00	Random Forest
27.3291	82.45	81.94	67.29	73.80	50.00	Gradient Boosted Decision Tree
27.3291	84.02	82.71	69.79	75.15	50.00	Recurrent Neural Network
27.3291	84.81	76.31	67.89	68.57	50.00	Baseline
27.3291	85.84	83.89	72.73	77.00	50.00	Elastic Net
27.3291	88.76	83.09	75.69	76.79	50.00	Neural Network
32.7215	78.54	76.99	54.63	65.59	60.00	Ideographic Decision Tree
32.7215	78.85	77.45	53.40	65.77	60.00	Ideographic Decision Tree Over 100 Days
32.7215	79.41	77.69	54.87	66.34	60.00	Ideographic Decision Tree Over 50 Days
32.7215	85.90	83.73	69.13	75.75	60.00	Random Forest
32.7215	86.53	76.05	66.83	67.36	60.00	Baseline
32.7215	88.76	83.49	73.24	76.19	60.00	Gradient Boosted Decision Tree
32.7215	89.80	85.19	75.62	78.47	60.00	Elastic Net
32.7215	90.78	82.20	75.20	75.15	60.00	Recurrent Neural Network
32.7215	91.35	80.73	75.08	73.62	60.00	Neural Network
38.1646	80.69	80.90	62.87	71.65	70.00	Baseline
38.1646	84.52	82.50	67.17	74.09	70.00	Ideographic Decision Tree
38.1646	84.64	83.78	69.06	75.96	70.00	Ideographic Decision Tree Over 50 Days
38.1646	87.35	85.71	74.69	79.23	70.00	Random Forest
38.1646	87.83	83.81	74.23	77.08	70.00	Gradient Boosted Decision Tree

Average number of days for training set size	Precision (%)	F_1 score (%)	Area under the curve	Accuracy (%)	Training set size (%)	Model
38.1646	88.57	84.21	75.36	77.72	70.00	Elastic Net
38.1646	88.60	88.99	76.76	83.25	70.00	Ideographic Decision Tree Over 100 Days
38.1646	89.16	80.69	74.07	73.88	70.00	Recurrent Neural Network
38.1646	89.57	82.76	75.69	76.28	70.00	Neural Network
43.2025	81.44	81.79	63.94	72.81	80.00	Baseline
43.2025	84.10	83.00	67.72	74.82	80.00	Ideographic Decision Tree
43.2025	88.31	90.07	76.80	84.62	80.00	Ideographic Decision Tree Over 50 Days
43.2025	89.00	84.20	75.63	77.71	80.00	Elastic Net
43.2025	89.00	83.43	75.18	76.80	80.00	Gradient Boosted Decision Tree
43.2025	89.25	84.70	76.22	78.35	80.00	Random Forest
43.2025	90.38	78.91	74.20	72.16	80.00	Recurrent Neural Network
43.2025	90.57	90.78	79.87	85.82	80.00	Ideographic Decision Tree Over 100 Days
43.2025	90.89	73.87	72.06	67.27	80.00	Neural Network

Predictors of Stress

In our ideographic models, there was significant heterogeneity in the effect of features of weather and exercise. Figure 2 shows a series of horizontal bar plots, visualizing the Gini importance values, or the factors that significantly predicted day-to-day variability in stress rating for the 9 individuals with at least 100 valid days of data.

The most frequent important predictors for each participant included daylight minutes, temperature, and exercise behavior for the current or preceding 3 days. Often, variables that were assessed closer (temporally) to the stress rating being predicted exhibited greater importance.

Viewing feature importance in aggregate for all the participants in the dataset, the Gini importance values from their corresponding decision tree model were sorted and then ranked. Thereafter, the number of times that each input variable ranked in the top 5 spots across participants' rankings was retrieved to create Figure 3. Decision tree models appear to have most frequently and highly ranked exercise-related features, followed by environmental variables.

Figure 4 depicts the relevance scores for the recurrent neural network model, which were derived using LRP. Like the decision tree models, the scores ranged widely, even among participants. To retrieve a metric more comparable with the strictly positive Gini importance values derived from the decision trees, the absolute value of the LRP scores was taken for each participant, followed by the median value for each input variable. The neural network model often ranked the highest, preceding stress self-reports, but there was also significant variety across features such as weather (eg, average or minimum temperature on the day of the stress rating), exercise (eg, number of minutes of exercise or number of exercise bouts the previous day), age, height, and weight.

In Figure 5, as with the decision tree models, the relevance of each input variable was ranked and then counted across participants.

Figure 6 depicts the F_1 score, AUC, and accuracy for the nomothetic recurrent neural network model and the ideographic decision trees across multiple training set sizes. For individuals with fewer than 50 valid days, the nomothetic models generally outperformed or performed comparably as the ideographic ones. Nomothetic model performance and ideographic model performance converged as training set size per participant increased. However, for the 16 individuals with more than 50 valid days and the 9 individuals with over 100 valid days, ideographic performance overtook nomothetic performance. Each dot corresponds to a training set size proportion, ranging from 30% to 80% in increments of 10%. The number of days included in the training set for each proportion across all participants, on an average, was as follows: 17 days for 30%, 22 days for 40%, 27 days for 50%, 33 days for 60%, 38 days for 70%, and 43 days for 80%.

Using LRP for the nomothetic model, a per-sample feature importance method, allows us to gain insight on what influenced a participant's predicted stress score for a given day. In Figure 7, a particular participant's actual stress self-reports, predicted stress self-reports, outside temperature, and the neural network model's LRP values are depicted across several consecutive days. On day 5, the neural network model assigned a high importance to the average temperature. As shown in the figure, as the average temperature increased (temperature here is standardized to fit on the same scale as the other values), predicted stress and the true stress value increased. Note that in the following graphs, the left-hand y-axis contains the scale for the input variable and the LRP value, and the right-hand y-axis contains the scale for both predicted and true stress rating. The x-axis simply denotes consecutive dates, marked as integers, rather than true dates.

Figure 2. Most important predictors for the ideographic decision tree models for the 9 individuals with at least 100 valid days of data.

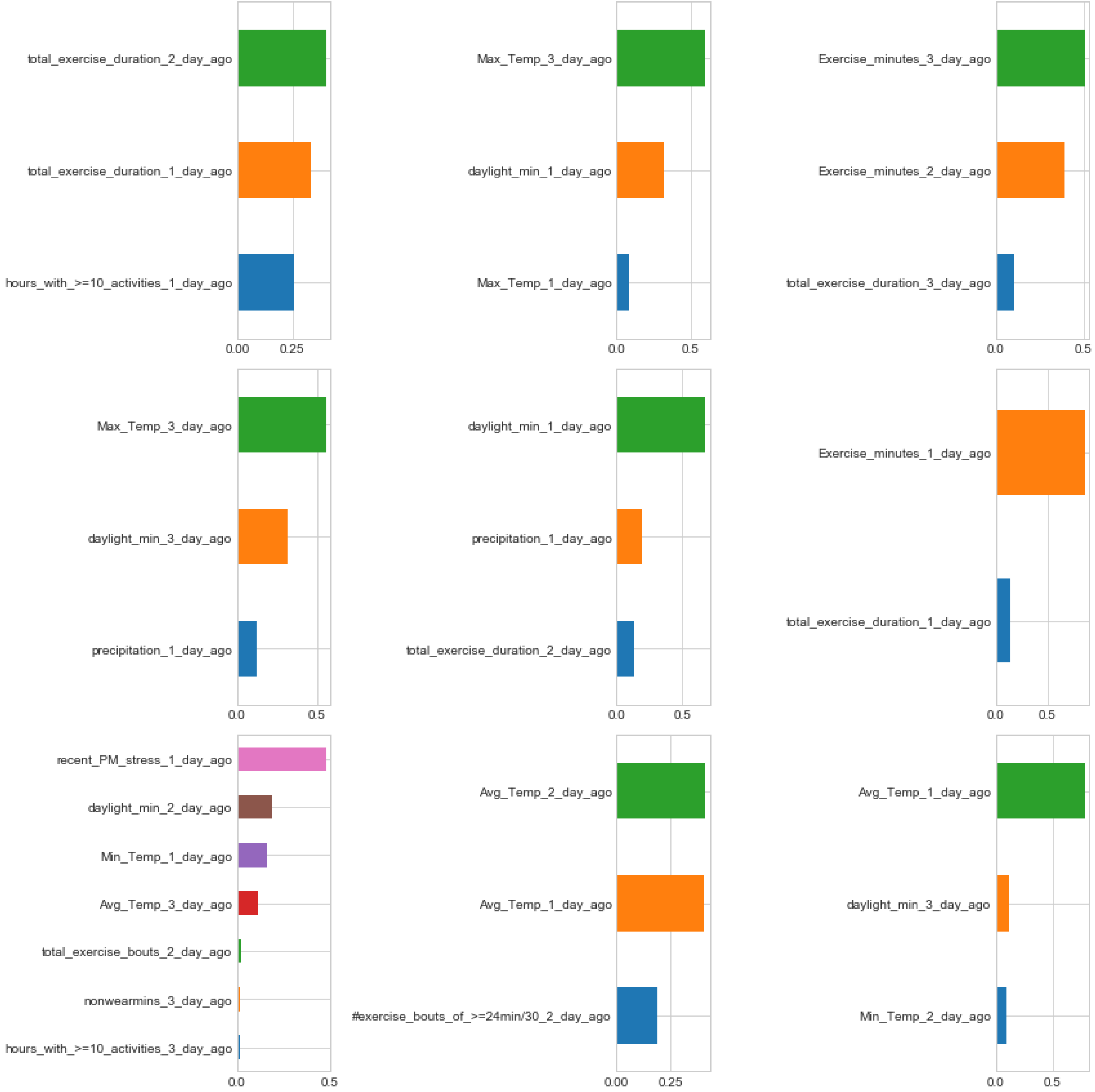


Figure 3. A horizontally stacked bar chart of occurrences of the most frequently appearing predictor variables, and how often they ranked in the top 5 spots across participants' predictor variable importance rankings from the ideographic models.

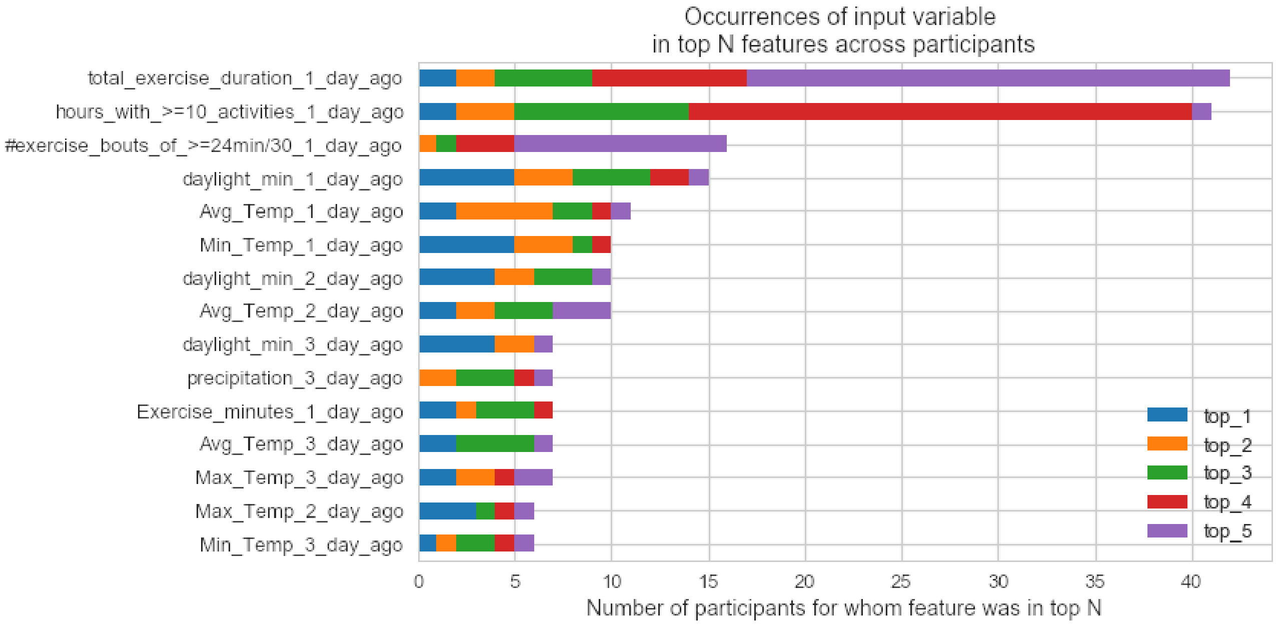


Figure 4. A horizontal bar chart of the 15 predictor variables with the highest median absolute layer-wise relevance propagation scores. LRP: layer-wise relevance propagation.

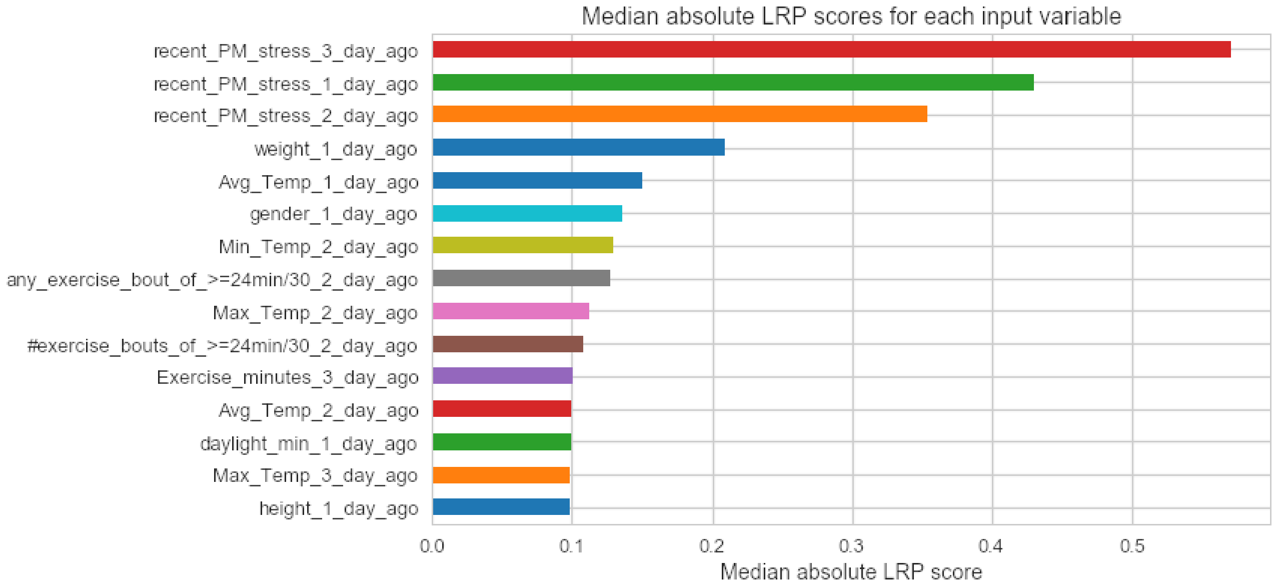


Figure 5. A horizontally stacked bar chart of occurrences of the most frequently appearing predictor variables, and how often they ranked in the top 5 spots across participants' predictor variable importance rankings from the nomothetic recurrent neural network model.

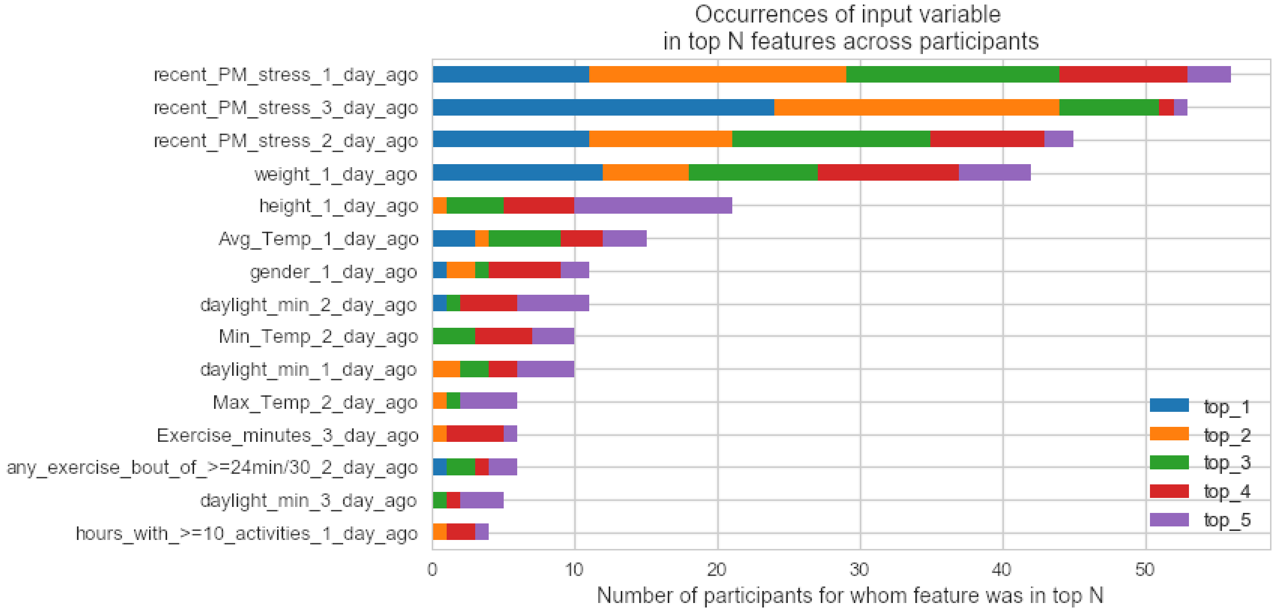


Figure 6. A plot comparing the F1 score, area under the curve, and accuracy across the neural network model and the ideographic models. Training set size varies on the x-axis, and performance of participants with more than 50 valid days and more than 100 valid days is shown separately for the ideographic models. The y-axis is the same in all 3 figures.

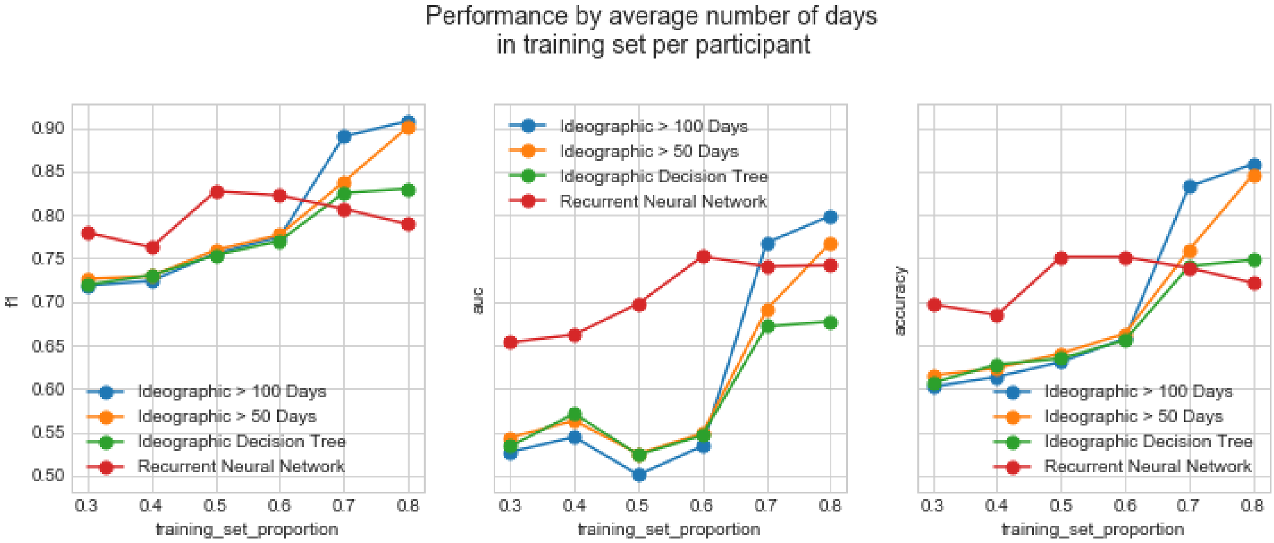


Figure 7. A plot of the actual stress rating, predicted stress rating, layer-wise relevance propagation value, and average temperature for an individual participant, with average temperature standardized to fit on the same graph. LRP: layer-wise relevance propagation.

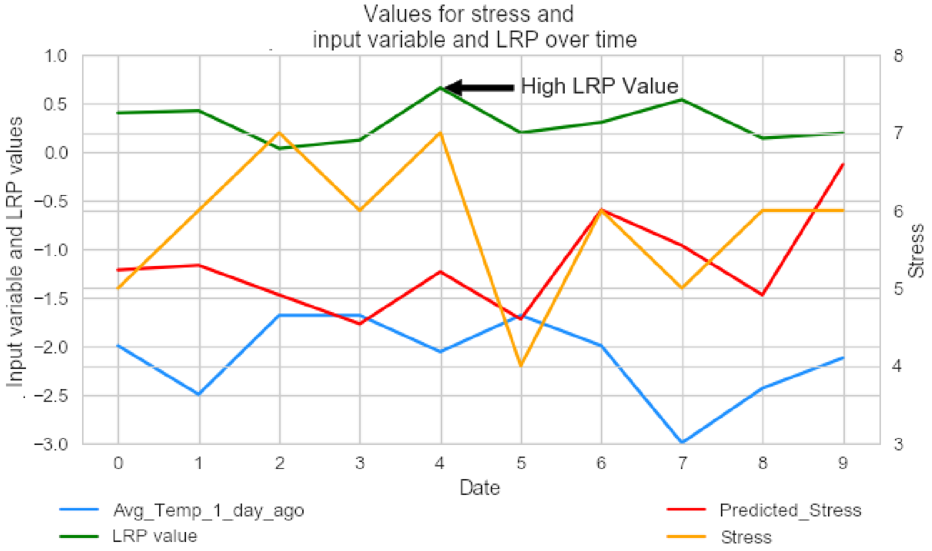
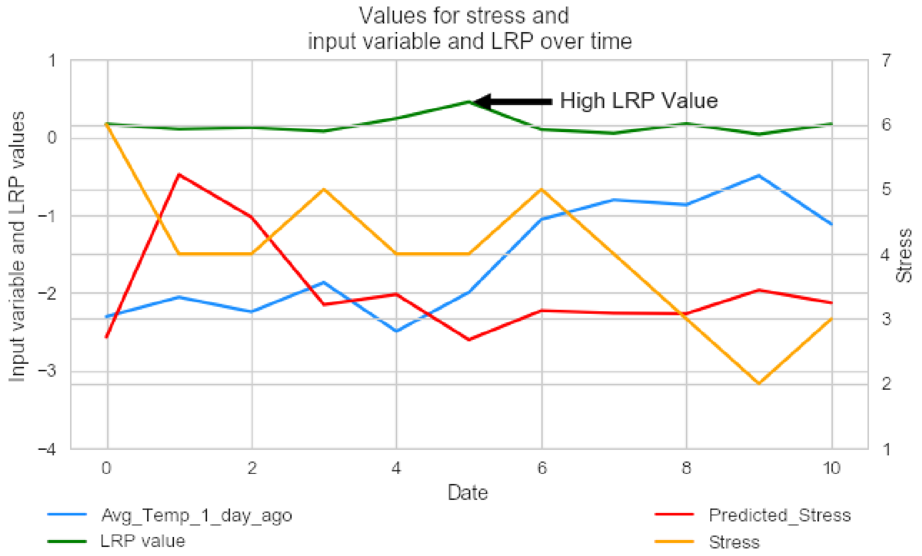


Figure 8. A plot of the actual stress rating, predicted stress rating, layer-wise relevance propagation value, and average temperature for an individual participant, with the average temperature standardized to fit on the same graph. LRP: layer-wise relevance propagation.



For another participant represented in Figure 8, an increase in the maximum temperature on a particular day was associated with a decrease in predicted stress, demonstrating individual-specific response patterns to environmental variables.

Exercise, as measured through total minutes or number of separate bouts, was often inversely related with stress rating, where less exercise increased the predicted stress rating (Figure 9).

However, there is a risk when interpreting feature importance presented as a time series. Although a variable may carry a high positive influence for a given day, it may be outweighed by a high cumulative negative impact of other variables, causing an improper inference. As such, it is useful to visualize a prediction for a given day not as a time series, but as a force plot of contributions from different variables, allowing us to more easily disentangle their influences. Treating LRP values as approximations for Shapley values, we can retrieve the following force plot depicted in Figure 10 of feature importance for a specific prediction.

Here, the predicted stress self-report value is 1.07. The stress rating 2 days ago of 2 and the stress rating 3 days ago of 6 push the prediction downward. The minimum temperature of 35°F 3 days ago and the absence of any exercise bouts of greater than 24 minutes of moderate-to-vigorous physical activity out of a 30-minute period push the prediction upward.

We can also aggregate these force plots across a set of predictions to understand how our nomothetic neural network model behaves more generally. In Figure 11, each horizontal line displays the LRP values for an input variable, across the entire testing set. The variables are sorted by the sum of the magnitudes of their values, in descending order. The figure shows that higher stress ratings preceding a predicted stress rating typically increase predicted stress, whereas lower minimum temperatures typically increase predicted stress.

Figure 9. A plot of the actual stress rating, predicted stress rating, layer-wise relevance propagation value, and exercise minutes for a participant, with exercise minutes standardized to fit on the same graph. A lower value for exercise minutes on the previous day was associated with a high layer-wise relevance propagation value and a higher predicted stress rating. LRP: layer-wise relevance propagation.

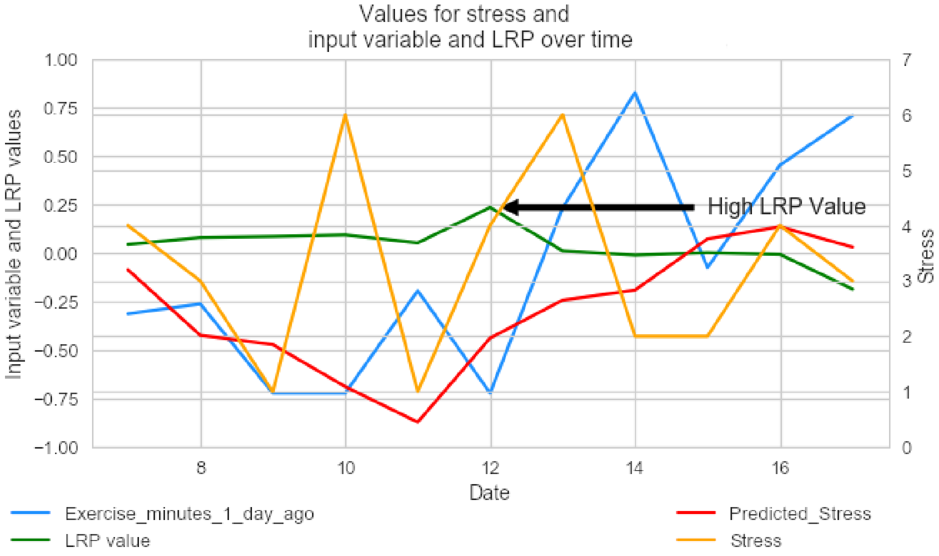


Figure 10. A force plot visualizing the contributions of different input variables to a predicted stress rating for a single participant.

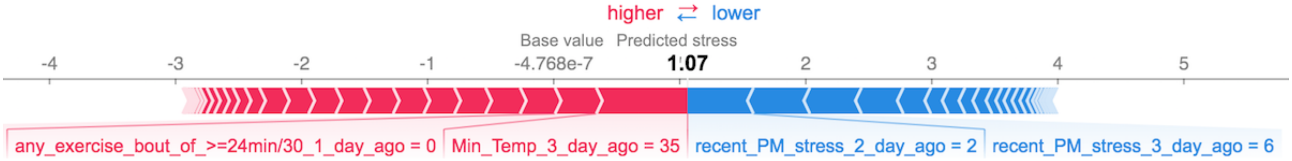
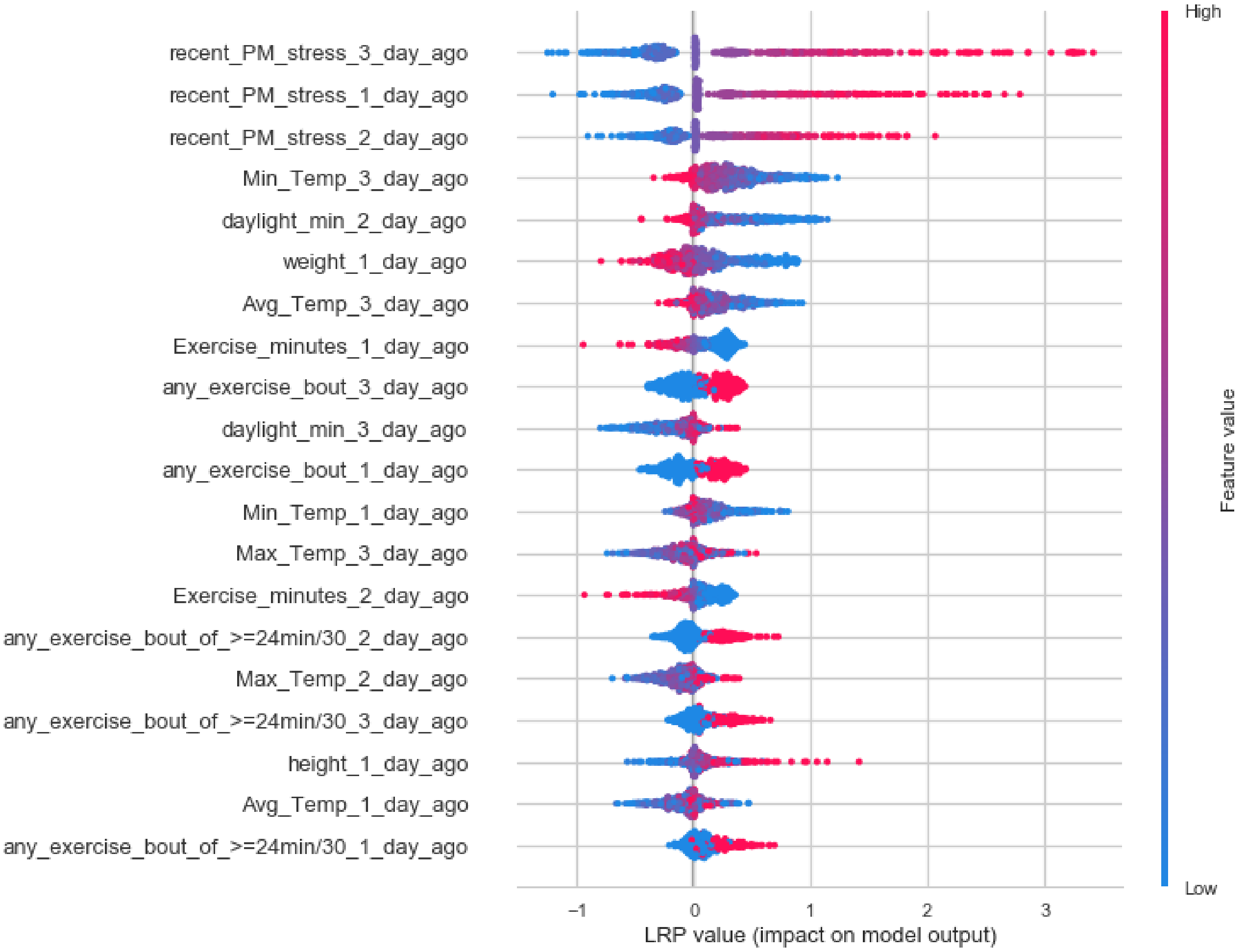


Figure 11. Layer-wise relevance propagation values for each input variable across the entire testing set. Each horizontal line corresponds to a single input variable. LRP: layer-wise relevance propagation.



Discussion

Overview

Many studies have examined relationships among exercise, weather, and stress using both linear and nonlinear approaches, and a mix of self-report questionnaire responses and automatically gathered sensing data. Some have taken an exclusively nomothetic approach; Wang et al [1] used both EMA data and automatically sensed activity and sociability data to explore correlates of stress, depression, and loneliness, but did not use a personalized machine learning approach or investigate individual predictors. Using a large set of meteorological data paired with responses to a self-report questionnaire and a mixed modeling approach, Beecher et al [2] found that increased sun exposure decreased reported distress.

In contrast, other studies have taken an ideographic approach. Tuarob et al [4] used a variety of machine learning techniques and questionnaire data to train ideographic models predicting participant mental states but relied on self-reported input data and did not investigate predictors for each individual. Sharmin et al [14] used sensor data and self-reported stress reports to create personalized visualizations that were then manually examined for temporal trends in stress. Plarre et al [5] trained ideographic decision trees using electrocardiographic and respiration-related data to predict self-reported stress after performing activities in the laboratory setting, such as public speaking or mental arithmetic. Burg et al [6] analyzed the same EMA and exercise data as those used here and estimated ideographic random coefficient mixed models; they found that the influence of exercise on self-reported stress was heterogeneous, as was the effect of self-reported stress on exercise.

Finally, taking an integrative approach and using the same dataset as that used by Burg et al [6] and us, Cheung et al [15] compared the performance of ideographic and nomothetic methods to predict whether an individual will exercise on a given day, again showing that for some, but not all participants, self-reported stress was a predictor of exercise.

In this paper, heterogeneity in the effects of predictors on stress was confirmed, highlighting the value of using an ideographic modeling approach. Further, it was demonstrated that the nomothetic model performs better (ie, is more accurate) than the ideographic model initially, but as data collected accumulates, the performance of the ideographic model equals and then surpasses that of the nomothetic model, providing a motivating example for a “warm start” strategy to leverage the advantages of each modeling approach. Put into practice, one might continuously monitor the performance of a nomothetic model and an ideographic model on a particular individual and adjust the weight of each model's contribution to the predicted output accordingly.

Retrieving individual predictors yields hypotheses that we can test for a particular participant and, perhaps, act on. For example, if a clinician repeatedly sees that low average temperatures are driving higher predicted stress values, they may be able to recommend a particular intervention for the patient. Of course, not all situations are so easily interpretable, and the relationship of a predictor to the outcome may not be linear. This is both an advantage and a disadvantage, allowing a model to potentially be flexible to multiple climates but limiting the ease of generalizability to other participants, and requires either manual inspection as mentioned above or a more rigorous and automated method of consistent linear predictor detection.

Here, the LRP values from our nomothetic neural network model, and Gini importance values from our ideographic decision trees, suggested different predictors for each individual. The nomothetic neural network model gave preceding stress self-reports higher weight on an average (Figure 10). This may be a result of training the neural network using dropout, in which increasing the weight given to preceding stress reports was an optimal way to manage the heterogeneity of participant response patterns. Notably, Sarker et al [16] reported similar results that stress episodes increase the likelihood of subsequent episodes, although these were within-day data.

Although Figure 4 presents a small sample of individuals and a relatively small dataset overall, it depicts the LRP values for the nomothetic neural network model and motivates discussion of real-world applications of such a model for just-in-time predictions. From this, it seems likely that it is effective to start with a nomothetic model to maximize predictive performance and gain initial insights into the possible correlates of stress before switching to (or increasing the weight of, with an ensemble-type implementation) an ideographic model once enough responses have been collected. This assumes that immediate interventions are necessary, valuable, and worth the cost of potentially introducing bias in the dataset for that individual. Alternatively, predictions could be withheld until sufficient data are collected for ideographic models to be used, but this might come at the expense of disengaging participants during the process of data collection. More generally, adopting a framework, as proposed by Nahum-Shani et al [17], to specify specific proximal outcomes while managing participant engagement is prudent.

Future Directions and Limitations

There is no current industry or academic standard for retrieving feature importance values from neural networks. Layer-wise relevance propagation, the method used here, has drawbacks of not meeting certain axiomatic properties of ideal feature importance methods [18]. In this case, however, LRP was used because it does not require a baseline, whereas other methods require some reference input for comparison. This instance is not one in which it is immediately clear what that input would be. Other model-agnostic, sensitivity-based approaches exist, such as Locally Interpretable Model Explanations [19]. Regardless of the approach for retrieving feature importance values, we retrieved per-participant predictors from the nomothetic neural network model by taking the median of the absolute value of LRP scores across a subset of the highest-stress events in a participant's test set. Other strategies may yield features that are more representative of an individual's stress. Further research in the field may be required to elaborate on strengths and weaknesses of different approaches in terms of interpretability, generalizability, and suitability for inclusion in a clinical decision process.

Next, in this setting, individuals who deviate significantly from the rest of the population may create large gradient updates to a neural network, potentially reducing performance for other individuals. Although this can be mitigated through techniques such as batch normalization and dropout, these strategies, in turn, reduce the degree to which the model can closely fit the data and obscure important differences among individuals. Similarly, individuals contributing a larger number of samples to the training set can bias the model. If an exhaustive hyperparameter search is performed without regularization strategies like L1/L2 penalties, max-norm constraints, dropout, or early stopping, the model may be further overfit to these individuals. As a result, some settings might still be best served by an ideographic approach for safety, but even N-of-1 decision trees may need to be constrained in their complexity to prevent overfit, especially while still acquiring data.

With self-report values, similar to item ratings, variable distributions may be skewed for particular individuals based on their perception of the scale. For example, in this dataset, some individuals never reported their stress to be above a value of 7, despite the scale going up to 10. Examples of these differing response patterns can be found in Figure 1. It is also highly unlikely that responses are missing at random. In fact, these days may be more stressful than the ones reported. The same issue may exist with Fitbit nonwear data. Time series forecasting methods often rely on either excluding or interpolating time windows that have missing data. Here, we do not interpolate missing data and instead, choose to exclude them. As a result, in situations in which self-report data are used, individual response patterns should be taken into account, whether through techniques such as feature engineering or increased data collection.

In our analyses, samples with missing stress ratings prior to the predicted stress rating were excluded. Based on the observed difference in distributions between stress ratings preceded by missing stress ratings, and ones that were not preceded by such ratings, it is possible that the ratings may not be missing at random. Instead, unobserved stress ratings may be indicative of higher stress in some patients. As a result, we primarily explored imputation using a dummy value, but models trained using these values severely underperformed, likely due to the relatively small size of the dataset; therefore, those results were not reported here. Alternatively, missing stress ratings could be imputed using a rule-based approach such as incrementing the most recently observed stress rating. However, for a practical implementation, we believe that with sufficient dataset size, missing stress ratings could be replaced by dummy values to avoid manually biasing predictions and improve generalization.

Further, current Fitbit devices and other accelerometers provide continuous heart rate monitoring data, which may provide additional useful predictors and mitigate the effect of missing stress ratings.

Finally, as confidence intervals were not retrieved, we lacked a measure of certainty per prediction. This could be alleviated by using Bayesian neural network or dropout-based methods.

Conclusions

Through the combination of a nomothetic neural network model, recent advances in retrieving per-sample feature importance, and ideographic decision trees, we show that high predictive performance can be achieved while recognizing individual differences and surfacing personalized predictors of stress. Key predictors in the nomothetic models were typically related to recent stress experience and weather activity. In addition, key predictors in the ideographic models displayed significant heterogeneity but were often weather or exercise related for individuals from whom more data were collected. Environmental variables were also shown to affect stress differently in different participants; for example, high temperatures predicted high stress in one individual but low stress in another. These predictors can be used to provide individuals with insights into what may contribute to their stress, as indicated by Yoon et al [20]. These models can also be operationalized to generate interventions or encouragements just before instances of high stress when the model predicts, with a sufficient degree of confidence, impending stress based on what is expected to be effective for that particular individual. Finally, ideographic models surpassed a nomothetic one after sufficient data collection, supporting the use of a “warm start” model-switching approach. Further work is needed to explore interpretable and repeatable ways to assess personalized predictors in nonlinear settings, as applied to disentangling correlates of stress.

Acknowledgments

This work was supported by funding (R01-HL115941) from the National Heart, Lung, and Blood Institute at the National Institutes of Health. IK was also supported by the National Center for Advancing Translational Sciences (UL1TR001873).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Performance metrics for each model.

[PDF File (Adobe PDF File), 90KB - jmir_v21i4e12910_app1.pdf]

References

1. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: UbiComp: ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 13-17, 2014; Seattle, WA. [doi: 10.1145/2632048.2632054]
2. Beecher ME, Eggett D, Erekson D, Rees LB, Bingham J, Klundt J, et al. Sunshine on my shoulders: Weather, pollution, and emotional distress. J Affect Disord 2016 Dec 15;205:234-238. [doi: 10.1016/j.jad.2016.07.021] [Medline: 27449556]
3. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
4. Tuarob S, Tucker CS, Kumara S, Giles CL, Pincus AL, Conroy DE, et al. How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information. J Biomed Inform 2017 Dec;68:1-19 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.010] [Medline: 28213145]
5. Plarre K, Raij A, Hossain S, Ali A, Nakajima M, Al-absi M, et al. Continuous inference of psychological stress from sensory measurements collected in the natural environment. 2011 Presented at: Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks; April 12-14, 2011; Chicago, IL.
6. Burg MM, Schwartz JE, Kronish IM, Diaz KM, Alcantara C, Duer-Hefele J, et al. Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional stress-exercise association at the group and person (N of 1) level. Ann Behav Med 2017 Dec;51(6):799-809 [FREE Full text] [doi: 10.1007/s12160-017-9902-4] [Medline: 28290065]
7. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S582-S588. [Medline: 16294121]
8. GitHub. 2018. Keras team URL: https://github.com/keras-team/keras [accessed 2018-11-27] [WebCite Cache ID 74ENlDUSZ]
9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine learning in Python. JMLR 2011;12:2825-2830.
10. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]

11. Bach S, Binder A, Montavon G, Klauschen F, Müller K, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015 Jul;10(7):e0130140 [FREE Full text] [doi: 10.1371/journal.pone.0130140] [Medline: 26161953]
12. Shrikumar A, Greenside P, Kundaje A. Learning Important Features Through Propagating Activation Differences. 2017 Presented at: Proceedings of the 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia p. 3145-3153.
13. Lundberg S, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: Neural Information Processing Systems; Dec 4-9, 2017; Long Beach, CA.
14. Sharmin M, Raij A, Epstien D, Nahum-Shani I, Beck J, Vhaduri S. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 07-11, 2015; Osaka, Japan.
15. Cheung Y, Hsueh P, Qian M, Yoon S, Meli L, Diaz KM, et al. Are Nomothetic or Ideographic Approaches Superior in Predicting Daily Exercise Behaviors? Methods Inf Med 2017;56(6):452-460 [FREE Full text] [doi: 10.3414/ME16-02-0051] [Medline: 29582914]
16. Sarker H, Hovsepian K, Chatterjee S. From markers to interventions: the case of just-in-time stress intervention. In: Mobile Health: Sensors, Analytic Methods, and Applications. New York: Springer; 2017:411-433.
17. Nahum-Shani S, Smith S, Tewari A, Witkiewitz K, Collins L, Spring B. Just-in-Time Adaptive Interventions (JITAIs): An Organizing Framework for Ongoing Health Behavior Support (Technical Report No. 14-126). University Park, PA: The Methodology Center; 2014.
18. Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for Deep Networks. In: ICML. 2017 Presented at: ICML; June 10-15, 2017; Long Beach, CA.
19. Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the predictions of any classifier. 2016 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939778]
20. Yoon S, Schwartz J, Burg M, Kronish IM, Alcantara C, Julian J, et al. Using Behavioral Analytics to Increase Exercise: A Randomized N-of-1 Study. Am J Prev Med 2018 Dec;54(4):559-567 [FREE Full text] [doi: 10.1016/j.amepre.2017.12.011] [Medline: 29429607]

Abbreviations

AUC: area under the curve
EMA: ecological momentary assessment
LRP: layer-wise relevance propagation

Edited by G Eysenbach; submitted 28.11.18; peer-reviewed by E Hekler, C Fernández; comments to author 03.01.19; revised version received 24.01.19; accepted 03.03.19; published 26.04.19.

Please cite as:
Rozet A, Kronish IM, Schwartz JE, Davidson KW
Using Machine Learning to Derive Just-In-Time and Personalized Predictors of Stress: Observational Study Bridging the Gap Between Nomothetic and Ideographic Approaches
J Med Internet Res 2019;21(4):e12910
URL: http://www.jmir.org/2019/4/e12910/
doi:10.2196/12910
PMID:31025942

Original Paper

How Confidence in Prior Attitudes, Social Tag Popularity, and Source Credibility Shape Confirmation Bias Toward Antidepressants and Psychotherapy in a Representative German Sample: Randomized Controlled Web-Based Study

Stefan Schweiger[1], Mag rer nat; Ulrike Cress[1], Dipl Psych, Dr rer soc

Knowledge Construction Lab, Knowledge Media Research Center, Leibniz-Institut für Wissensmedien, Tuebingen, Germany

Corresponding Author:
Stefan Schweiger, Mag rer nat
Knowledge Construction Lab
Knowledge Media Research Center
Leibniz-Institut für Wissensmedien
Schleichstrasse 6
Tuebingen, 72076
Germany
Phone: 49 6702023838
Email: schw.stefan@gmail.com

Abstract

Background: In health-related, Web-based information search, people should select information in line with expert (vs nonexpert) information, independent of their prior attitudes and consequent confirmation bias.

Objective: This study aimed to investigate confirmation bias in mental health–related information search, particularly (1) if high confidence worsens confirmation bias, (2) if social tags eliminate the influence of prior attitudes, and (3) if people successfully distinguish high and low source credibility.

Methods: In total, 520 participants of a representative sample of the German Web-based population were recruited via a panel company. Among them, 48.1% (250/520) participants completed the fully automated study. Participants provided *prior attitudes* about antidepressants and psychotherapy. We manipulated (1) *confidence* in prior attitudes when participants searched for blog posts about the treatment of depression, (2) *tag popularity* —either psychotherapy or antidepressant tags were more popular, and (3) *source credibility* with banners indicating high or low expertise of the tagging community. We measured *tag* and *blog post* selection, and *treatmentefficacy ratings* after navigation.

Results: Tag popularity predicted the proportion of selected antidepressant tags (beta=.44, SE 0.11; *P*<.001) and blog posts (beta=.46, SE 0.11; *P*<.001). When confidence was low (−1 SD), participants selected more blog posts consistent with prior attitudes (beta=−.26, SE 0.05; *P*<.001). Moreover, when confidence was low (−1 SD) and source credibility was high (+1 SD), the efficacy ratings of attitude-consistent treatments increased (beta=.34, SE 0.13; *P*=.01).

Conclusions: We found correlational support for defense motivation account underlying confirmation bias in the mental health–related search context. That is, participants tended to select information that supported their prior attitudes, which is not in line with the current scientific evidence. Implications for presenting persuasive Web-based information are also discussed.

Trial Registration: ClinicalTrials.gov NCT03899168; https://clinicaltrials.gov/ct2/show/NCT03899168 (Archived by WebCite at http://www.webcitation.org/77Nyot3Do)

(J Med Internet Res 2019;21(4):e11081) doi:10.2196/11081

KEYWORDS

attitude; psychotherapy; antidepressive agents; culture; Germany; health literacy; professional competence; information systems; consumer health information; information dissemination

Introduction

Background

Do people attend to information independent of their prior attitudes, and do they distinguish expert from nonexpert sources on the Web? To address these important questions [1-3], we investigate confirmation bias, the tendency to favorably select and evaluate attitude-consistent information [3-6].

A comprehensive meta-analysis identified 2 major motivational factors that moderate confirmation bias [7]. First, when we face information that suggests our point of view is wrong, we try to maintain our prior attitudes by choosing and believing in attitude-consistent information, which is called *defense motivation* [7-9].

In contrast to this, in some situations, we may be genuinely interested in acquiring objectively correct and accurate information [7,8,10]. This *accuracy motivation* can guide our information search, even when information is not consistent with our prior attitudes [7]. Particularly in the health context, we should form attitudes independent of our defense mechanisms and base evaluations on objectively correct information. In the following sections, we outline 3 factors that may reduce confirmation bias, given that we are accuracy motivated when searching for mental health–related information.

Confidence and Confirmation Bias

First, low confidence should decrease confirmation bias [7]. However, people tend to be overly confident in prior attitudes and knowledge [11,12] in a large range of domains, such as academic, intellectual, vocational, athletic, and medicine [13]. When people are overly confident in their prior attitudes, confirmation bias increases [14].

For the mental health–related context, it is important that confidence varies for people with different mental disorders [15]. For example, individuals who experience anxious and depressive symptoms show less than average confidence (but average accuracy) in decision-making tasks [15], which suggests that they could be even less prone to confirmation bias.

Looking at how to influence confidence, overconfidence can be reduced when participants reflect on their ability to describe, in a step-by-step manner, the causal functioning of objects to experts [16]. We draw on a manipulation that focused on people recalling situations where they were either confident or doubtful about their own thoughts—study 3 [17]. When participants recalled situations in which they had been confident (vs doubtful), and subsequently provided arguments about a controversial topic, they were more (or less) confident about their arguments [17,18]. For this study, one main goal was to replicate the manipulation (study 3 in [17]) with a representative sample, in the mental health context.

A recent review has shown that confidence manipulations tend to increase confirmation bias, which is explained by the defense motivation account [7]. According to defense motivation, when people have low confidence, they aim to defend their self-concept by selecting information that is in line with their attitudes. In contrast to this, we draw on a metacognitive manipulation of confidence that aims to make people perceive their current thoughts as less valid—study 3 in [17]. Consequently, they should perceive their attitudes as less valid (independent of their self-concept), and confirmation bias should decrease, given that searchers aim for valid information.

We expect that when prior attitudes are held with high confidence, participants preferably select and evaluate attitude-consistent information. If participants were defense motivated, high (vs low) confidence would make them less (vs more) threatened by attitude-inconsistent information and they would select more attitude-inconsistent information and evaluate it more favorably [7].

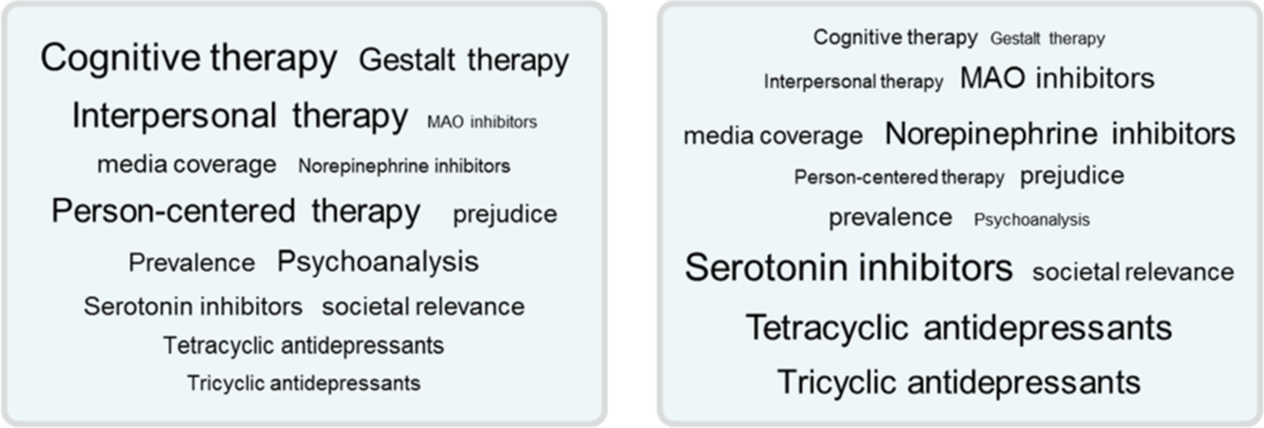
Social Tags as Signposts to Popular Information

The second influence on confirmation bias occurs when people face cues from socially aggregated information on the Web [19-24]. Cues indicating socially aggregated information include star ratings, likes, retweet counts, or social tags. In the case of tagging, tag clouds arise when users label or tag content on the Web, such as videos, images, or documents (Figure 1) [25,26]. When tags from the tagging community are aggregated and presented in tag clouds, the tags represent the consent of a majority of people and guide information searchers [19,20]. High majority consent or high tag popularity translates into large tags, which attract more attention than smaller tags with less social consent.

We suggest that social tag clouds are particularly nonintrusive and therefore highly suited to circumvent the influence of prior attitudes as larger tags are visually dominating, and it has been shown that people who primarily attend to large tags [25,27,28] are more likely to click on large tags [20,29,30] even when large tags are inconsistent with activated associations in memory [29,30] or prior attitudes [20]. Moreover, social consent elicits the behavior that conforms to the majority in offline settings [31,32].

Moreover, people select more trustworthy results when facing a grid-like (vs list-like) arrangement of search results, similar to social tag clouds [33]. In sum, tag clouds should be suited to decrease the influence of prior attitudes in information search and reduce confirmation bias.

Figure 1. The tag clouds used in the present study. Either psychotherapy (left), or antidepressants (right) were more popular.



Source Credibility of the Tagging Community

People do not always successfully consider high-quality information [34,35], particularly when browsing user-generated content [34]. A total of 2 meta-analyses concluded that personal characteristics [35], as well as platform characteristics [34], play an important role. The relationship between manipulated source credibility and perceived information credibility is higher for student samples (vs nonstudent samples) [35], and for user-generated content that is presented on common websites (vs blogs and discussion boards) [34].

Besides education, epistemic beliefs can influence how people perceive source credibility. For example, when searching information on 2 competing therapies for Bechterew disease, not all participants took source credibility into account [36]. Particularly, participants who viewed the Web as a reliable (vs unreliable) source of accurate knowledge did not reflect upon source credibility, they viewed URLs for a shorter time and selected less search results at the bottom of the page.

To our knowledge, there is a single study that uses tag clouds to investigate source credibility in the health context [37]. People searched for information on how to manage diabetes on a health forum with tag clouds [37]. In the first task, they searched for information that was of general interest, and in a second task, they searched for credible information. The tag cloud comprised 12 filler tags (eg, glucose, diet, and exercise), and 6 tags indicating source credibility of content (author, date, quote, reference, statistics, and testimonial). When participants browsed for general interest, only one-third used at least a source credibility tag. When explicitly asked to take source credibility into account, 90% used at least 1 source credibility tag.

It remains an open question whether people in a representative sample take the source credibility on a social tagging platform into account. In line with the accuracy motivation account, we expect that if information searchers recognize high source credibility, they will select more tags and related blog posts in total, regardless of whether attitude-consistent or attitude-inconsistent tags are more popular in the social tag cloud. If, on the other hand, people showed defense motivation [7], they would avoid attitude-inconsistent tags and blog posts with high source credibility and evaluate it less favorably.

Prior Attitudes Toward Antidepressants and Psychotherapy

With respect to the treatment of depressive disorders, people clearly favor psychotherapy over antidepressants [38-44]. Attitudes of laypeople manifest in estimated treatment efficacy as well as treatment recommendations [20,41,43,45,46]. People believe antidepressants to be little to moderately effective, whereas psychotherapy is believed to be moderate to highly effective [20]. As literature shows about equal, moderate efficacy of both types of treatment [47-49], people's attitudes and recommendations are biased.

We expect more positive prior attitudes toward psychotherapy than toward antidepressants in the German population, and with this study, we aim to describe the magnitude of the psychotherapy preference and present the arguments that shape these biased attitudes.

Hypotheses

First, we expect that people's attitudes (H1a) and efficacy ratings (H1b) before navigation are more favorable for psychotherapy than for antidepressants.

We expect to replicate Study 3 in [17]: after recalling situations in which participants were confident (vs doubtful), they should be more confident in their own arguments (H2a). We expect that high (vs low) *confidence* leads to a more pronounced confirmation bias and an increased selection of attitude-consistent tags (H2b) and blog posts (H2c), and this will strengthen the attitudes people already had before navigation (H2d). So, when prior attitudes favor psychotherapy, and confidence is high, participants prefer psychotherapy tags and blog posts and change their attitudes even more toward psychotherapy. If confidence is low, prior attitudes should not be related to selection of tags and blog posts and attitude change.

Tag popularity should circumvent the influence of prior attitudes, so participants select popular tags more frequently than less popular tags (H3a) and blog posts (H3b). Consequently, attitudes change in line with tag popularity (H3c).

Participants distinguish high from low *source credibility* (H4a). When tags and blog posts are collected by experts (vs novices), participants click on more tags (H4b) and blog posts (H4c) overall, independent of their prior attitudes, and people should show more attitude change for both treatments (H4d).

Methods

Participants

A representative sample with respect to age and gender was randomly drawn from a pool of a panel company. In total, 520 participants started the fully automated Web-based study, 48.1% (250/520) completed it, 1.3% (7/520) withdrew their data, and 3.2% (17/520) participants were dropped as they did not provide responses (Figure 2). Age of the remaining 43.5% (226/520) participants ranged from 18 to 60 years (mean 40.36, SD 12.17), and 50.0% (113/226) were female (Table 1). With respect to familiarity of the technology used in the study, 24.8% (56/226) stated they were familiar with the term *tag cloud*, 36.7% (83/226) stated they had already clicked on single tags to navigate the Web. Ethical approval was granted by the Ethical Committee of the Knowledge Media Research Center (LEK 2014/006).

Figure 2. CONSORT flow diagram. AD: antidepressants; PT: psychotherapy.

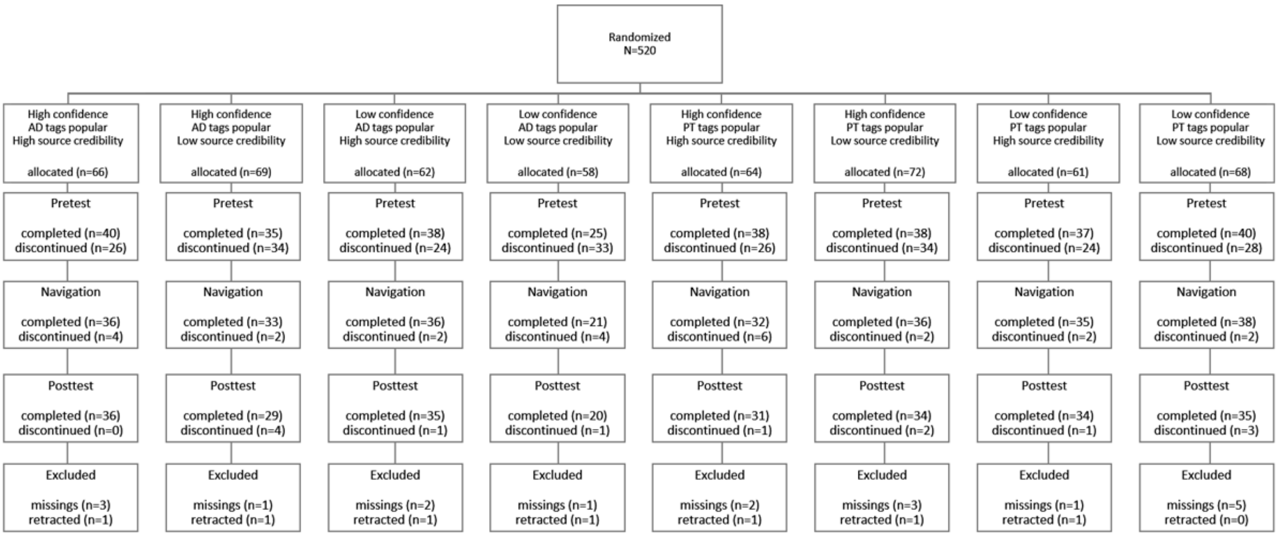


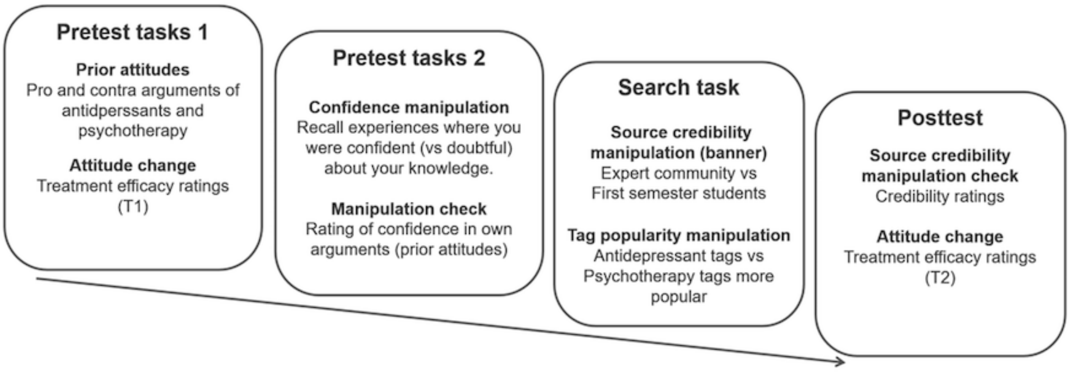
Table 1. Sample characteristics (N=226).

Characteristic	Statistics, n (%)
Education	
Qualified job	18 (8.0)
Abitur certificate	102 (45.1)
University degree	53 (23.5)
Other	53 (23.5)
Age in years	
18-19	9 (4.0)
20-29	46 (20.4)
30-39	45 (19.9)
40-49	65 (28.8)
50-59	52 (23.0)
60	9 (4.0)
Gender	
Male	113 (50.0)
Female	113 (50.0)

Procedure and Design

This study comprised a 2 (confidence: high and low) × 2 (tag popularity: antidepressants high and psychotherapy high) × 2 (tagging source credibility: high and low) between-subjects design. Participants enrolled via a Web-based portal of a private panel company (respondi AG, Cologne, Germany; ISO 26362 certified), which linked to our survey, and participants were offered €4 to complete it. First, participants were welcomed and informed that they could withdraw participation at any point. Participants were granted anonymity and asked to provide informed consent by clicking the button to start the study, after which they were randomly assigned to 1 of the 6 experimental conditions by a computerized random number procedure. Then, for *prior attitudes*, we asked participants to state pro and contra arguments regarding antidepressants and psychotherapy (pretest tasks 1; Figure 3). Next, they rated the efficacy of antidepressants and psychotherapy on scales. Then, they provided responses for an allegedly unrelated pilot study, which served to manipulate *confidence* [17]. Participants were asked to recall situations in which they had felt either confident or doubtful about their own knowledge (Study 3 in [17]). After this, they were asked to think back about their arguments regarding psychotherapy and antidepressants and they rated how confident they were about the arguments they had provided before. This rating served as a manipulation check for confidence. Next, participants searched for information about treatment efficacy to provide treatment advice for a hypothetical, closely-related person. To manipulate *source credibility*, we informed them that the blog post and responding tag had been gathered by a community of either experts in the field, such as experienced psychiatrists and by psychotherapists (high source credibility condition), or by psychology students and medical students in their first semester (low source credibility condition). To manipulate *tag popularity*, either psychotherapy or antidepressant tags were larger (Figure 1). They could also provide tags for blog posts. After 5 min of browsing in the tagging environment, a *Next* button appeared and from then on, participants could decide when to stop browsing tags and related blog posts. After navigation, participants rated source credibility (manipulation check) and provided efficacy ratings again. At the end of the study, participants could provide feedback in a text box.

Figure 3. Experimental procedure.



Materials

For the information search task, we provided a tagging environment (Multimedia Appendix 1). We presented 14 tags on the right side of the screen in which 5 tags represented psychotherapy, 5 tags represented antidepressant treatments, and 4 filler tags were irrelevant for treatment (prejudice, media coverage, societal relevance, and prevalence). Depending on the experimental condition, either psychotherapy-related tags or antidepressants-related tags were larger (ie, more popular). When participants clicked on a tag, 3 blog posts were presented on the left (Multimedia Appendix 1). Each blog post described a symptom of depressive disorders and the scientific studies on the efficacy of the treatment. In a pilot study, we had assured that the blog posts had equal persuasiveness. First, only the headline and the first sentence of each of the 3 related blog posts were shown. To read the full post, participants clicked on *(more...)*.

Independent Variables

Prior Attitudes

As index of pro and contra arguments for psychotherapy and antidepressants, we subtracted the number of arguments favoring antidepressants (contrapsychotherapy and proantidepressants) from the number of arguments favoring psychotherapy (propsychotherapy and contraantidepressants). Positive values of this index thus indicate a preference for psychotherapy. Arguments were rated by 2 raters (r=.78; P<.001), where differences were resolved by agreement.

Confidence

We adapted the experimental procedure by Petty and colleagues (Study 3 in [17]) and participants recalled situations in which they had felt confident or doubtful about their own knowledge, using 5 input text boxes for 5 min.

Tag Popularity

For the psychotherapy popular group, psychotherapy tags were larger, and for the antidepressant popular group, antidepressants tags were larger (Figure 1).

Source Credibility

On top of the page, banners showed that either alleged college students (low source credibility; Figure 4) or domain experts (high source credibility; Figure 5) had collected and tagged the blog posts. After the search task, participants rated the source credibility of the information on a scale from 1 (not at all) to 7 (highly).

Figure 4. Banner for the low source credibility condition.

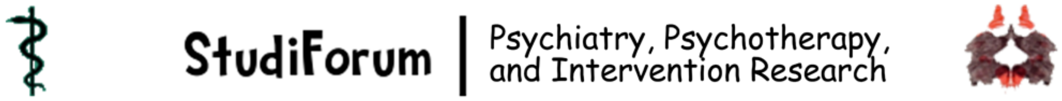


Figure 5. Banner for the high source credibility condition.

Confidence Ratings (Manipulation Check)

After participants listed situations in which they had been confident or unconfident, they rated confidence in their own arguments regarding prior attitudes on a scale from 1 (not at all) to 7 (highly). They were asked how the following words described their arguments: obvious, dubious, justified, credible, factual, well-founded, persuasive, and objective (Cronbach alpha=.88).

Source Credibility Ratings (Manipulation Check)

Participants rated the degree to which the following words described the tagging community: informed and competent (r=.70; P<.001).

Dependent Variables

Efficacy Ratings (Attitude Change)

Participants agreed to statements on the efficacy of psychotherapy and antidepressants on a scale from 1 (completely disagree) to 7 (completely agree), before (antidepressants Cronbach alpha=.89 and psychotherapy Cronbach alpha=.92) and after navigation (antidepressants Cronbach alpha=.94 and psychotherapy Cronbach alpha=.95). To predict attitude change with respect to treatment preference, we derived a difference index score, subtracting the antidepressant from psychotherapy treatment ratings.

Beside attitude change in terms of treatment preference, we analyzed pooled attitude change by taking the sum of efficacy ratings for both treatments before and after navigation (divided it by the number of items for interpretability).

Tag and Blog Post Selection

To measure attitude-consistent navigation, we recorded the number of tags and blog posts selected for each treatment category (0=psychotherapy and 1=antidepressants).

Results

All analyses presented were conducted with the R Software (R Foundation for Statistical Computing; Version 3.3.4); raw data and the analysis script can be found in Multimedia Appendix 2.

Prior Attitudes

As expected in H1a, we found that participants' prior attitudes favor psychotherapy over antidepressants. Participants stated more arguments for psychotherapy (mean 1.69, SD 1.77) than for antidepressants (mean 1.06, SD 1.51; t_{225}=5.30; P<.001, d=0.26), and they stated more arguments against antidepressants (mean 1.51, SD 1.53) than against psychotherapy (mean 0.73, SD 1.54; t_{225}=8.13; P<.001, d=0.34). We also descriptively analyzed arguments and pooled them into qualitative categories (Figure 6).

With H1b, we expected that people would provide more favorable efficacy ratings for psychotherapy compared with antidepressants before navigation. Participants rated statements about the efficacy of both treatments on 8 items, on a scale from 1 to 7 (Figure 7). As the internal consistency was high for both scales (antidepressants Cronbach alpha=.89 and psychotherapy Cronbach alpha=.92), we pooled them. A paired t test showed a moderate effect on the preference for psychotherapy (mean 5.24, SD 1.10) over antidepressants (mean 4.61, SD 1.19; t_{225}=9.71; P<.001, d=0.56; see items and response distribution in Figure 7). In sum, prior attitudes measured via pro and contra arguments, as well as via efficacy ratings, favored psychotherapy over antidepressants. Both measures were moderately correlated (r=.41; P<.001).

Figure 6. Arguments for and against the 2 treatments.

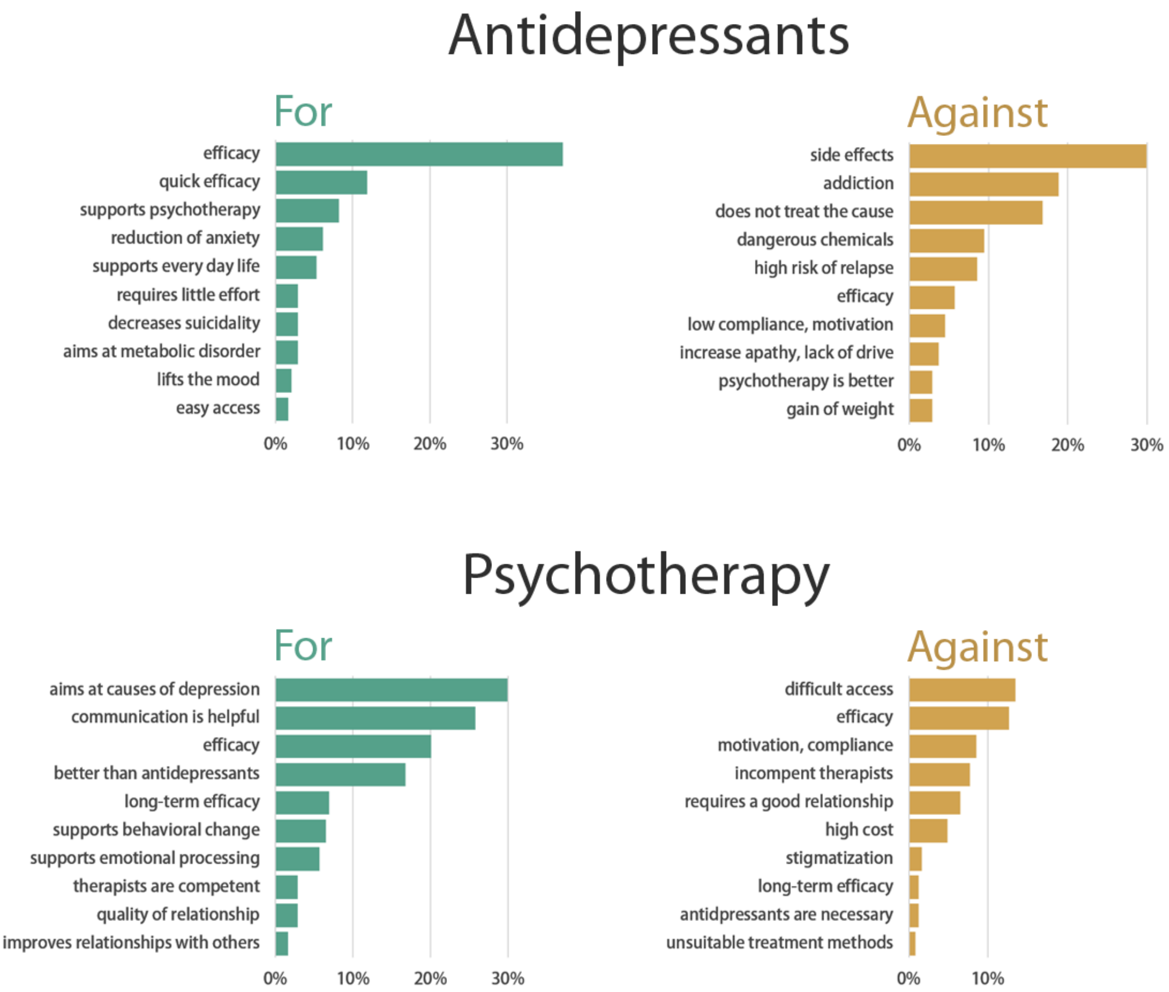
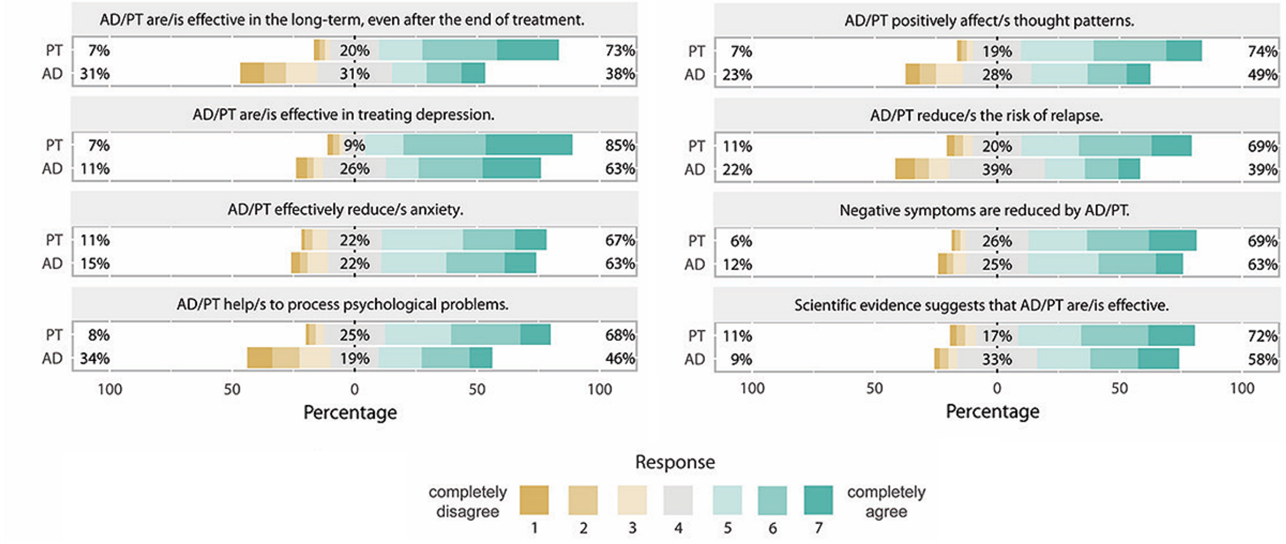


Figure 7. Prior attitudes about psychotherapy and antidepressants before information search. AD: antidepressants; PT: psychotherapy.



Manipulation Checks

Confidence

Contrary to our expectations (H2a), we could not replicate the confidence manipulation (Study 3 in [17]). After recalling situations in which they had been confident (mean 4.64, SD 1.20), participants were not more confident about their arguments compared with recalling situations in which they had been doubtful (mean 4.68, SD 1.09; t_{224}=<1; P=.78).

Source Credibility

In contrast to our expectation (H4a), source credibility ratings in the high source credibility condition (mean 4.87, SD 1.26) did not significantly differ from source credibility ratings in the low source credibility condition (mean 5.16, SD 1.32; t_{224}=1.67; P=.10).

As the confidence and source credibility manipulations were ineffective, we used respective manipulation check scores in the following regression analyses as predictors.

Confidence in Prior Attitudes

Tag Selection

To analyze attitude-consistent and attitude-inconsistent tag selection, we conducted logistic regressions with the dependent variable clicks on treatment tags. The number of clicks on the respective treatment (0=psychotherapy tag selected and 1=antidepressant tag selected) was entered in a logistic regression (Table 2). As predictors, we entered *prior attitudes* and *tag popularity* (0=psychotherapy tags popular and 1=antidepressant tags popular), *confidence ratings*, and *source credibility ratings* (see independent variables). We included 2-way interaction terms (Step 2 in Table 2) and tested for interactions with likelihood ratio tests [50,51].

Table 2. Selection ratio of antidepressant tags.

Predictor	Step 1[a]			Step 2[b]		
	Beta[c]	SE	*P* value	Beta	SE	*P* value
Intercept	−.39	0.08	<.001	−.41	0.08	<.001
Prior attitudes	−.02	0.03	.37	−.03	0.03	.32
Confidence score	.002	0.05	.97	.04	0.05	.82
Tag popularity	.44	0.11	<.001	.44	0.11	<.001
Source credibility score	−.005	0.04	.92	−.005	0.04	.92
PA[d]×confidence score	—[e]	—	—	−.01	0.03	.65

[a] Model fit: χ^2_4=17.1; P=.002 (Step 1).

[b] Model fit change: χ^2_1=0.2; P=.65 (vs Step 2).

[c] Continuous predictors were centered.

[d] PA: prior attitudes.

[e] Interaction term not included.

We expected that high confidence should strengthen the relationship between prior attitudes and the proportion of clicks on attitude-consistent tags (H2b). However, there was no significant interaction of the predictors' confidence in prior attitude ratings and prior attitudes predicting the selection of antidepressant tags (Step 2 in Table 2). As likelihood ratio tests showed, including 3-way interaction (χ^2_{10}=4.9; P=.90) and 4-way interaction (χ^2_{11}=5.0; P=.93) did not improve model fit.

Blog Post Selection

A second logistic regression used the same predictors as in the regression predicting tag selection but with blog post selection as criterion variable (Table 3). We expected that high confidence should strengthen the impact of prior attitudes and consequently lead to increased proportion of clicks on attitude-consistent blog posts (H2c). We observed an interaction between confidence and prior attitudes (beta=.11, SE 0.02; P<.001). To disentangle the interaction, we compared slopes of high (+1 SD) and low (−1 SD) confidence ratings. This showed that when confidence ratings were low (−1 SD), participants selected a higher proportion of blog posts that were in line with their prior attitudes (beta=−.26, SE 0.05; P<.001; Figure 8). When confidence ratings were high (+1 SD), there was no association with prior attitudes (beta=.02, SE 0.03; P=.57; Figure 8). In contrast to our expectation, and in line with the defense motivation account, when confidence was low but not high, there was an association between prior attitudes and selection of attitude-consistent blog posts.

Compared with the model including the 2-way interaction term (Step 2 in Table 3), neither including 3-way interaction term (χ^2_4=5.8; P=.21) nor including the 4-way interaction term (χ^2_5=6.0; P=.31) yielded a better model fit (all respective lower-order interaction terms were included as well).

Table 3. Selection ratio of antidepressant blog posts.

Predictor	Step 1[a]			Step 2[b]		
	Beta[c]	SE	*P*	Beta	SE	*P*
Intercept	−.75	0.07	<.001	−.87	0.08	<.001
Prior attitudes	−.05	0.03	.06	−.12	0.03	<.001
Confidence score	−.11	0.05	.02	−.04	0.05	.46
Tag popularity	.44	0.11	<.001	.45	0.11	<.001
Source credibility score	.02	0.04	.73	.03	0.04	.52
PA[d]×confidence score	—[e]	—	—	.11	0.02	<.001

[a] Model fit: χ^2_4=30.4; *P*<.001.

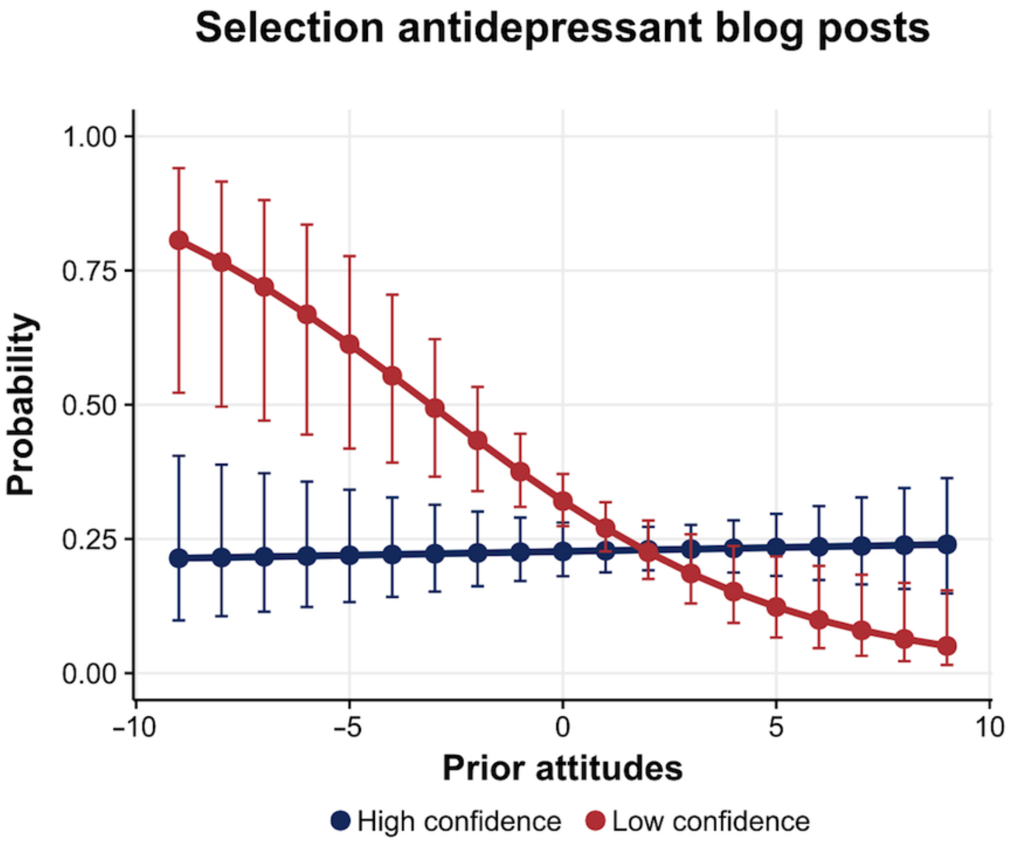
[b] Model fit change (vs Step 1): χ^2_1=25.6; *P*<.001.

[c] Continuous predictors were centered.

[d] PA: prior attitudes.

[e] Interaction term not included.

Figure 8. Predicted proportion of antidepressant blog posts selected, for high (+1 SD) and low (−1 SD) confidence (95% CI), with negative values indicating a preference for antidepressants over psychotherapy.



Attitude Change

We conducted multiple linear regressions. First, with the predictor variables prior attitudes, confidence ratings, and source credibility ratings (all centered), and the dichotomous variable tag popularity (0=psychotherapy popular and 1=antidepressants popular). In addition, we included the predictor difference score of efficacy ratings (antidepressants subtracted from psychotherapy) before navigation to analyze attitude change with a covariate approach [52]. As a criterion for attitudes after navigation, we included the difference score of efficacy ratings (Table 4).

Table 4. Treatment efficacy ratings (psychotherapy-antidepressants) after navigation.

Predictor	Step 1[a]			Step 2[b]			Step 3[c]		
	Beta[d]	SE	*P* value	Beta	SE	*P* value	Beta	SE	*P* value
Intercept	.70	0.08	<.001	.73	0.08	<.001	.74	0.08	<.001
Efficacy ratings (PT[e]−AD[f]) before navigation	.79	0.06	<.001	.78	0.06	<.001	.79	0.06	<.001
Tag popularity	−.05	0.12	.64	−.06	0.12	.60	−.05	0.11	.65
PA[g]	.08	0.03	.01	.10	0.06	.005	.12	0.03	<.001
Confidence score	−.08	0.05	.16	−.10	0.06	.08	−.10	0.06	.07
Source credibility score	.04	0.05	.34	.05	0.05	.33	.11	0.05	.03
PA×confidence score	—[h]	—	—	−.04	0.03	.16	−.07	0.03	.02
PA×source credibility score	—	—	—	—	—	—	.02	0.02	.40
Confidence score×source credibility score	—	—	—	—	—	—	−.03	0.04	.53
PA×confidence score×source credibility score	—	—	—	—	—	—	−.07	0.02	<.001

[a] Model fit: adjusted R^2=0.49, $F_{5,220}$=43.68; *P*<.001.

[b] Model fit (vs Step 1): Δ adjusted R^2<0.01, $F_{1,219}$=1.94; *P*=.16.

[c] Model fit (vs Step 2): Δ adjusted R^2=0.02, $F_{3,216}$=3.88; *P*=.01.

[d] Continuous predictors were centered.

[e] AD: antidepressants.

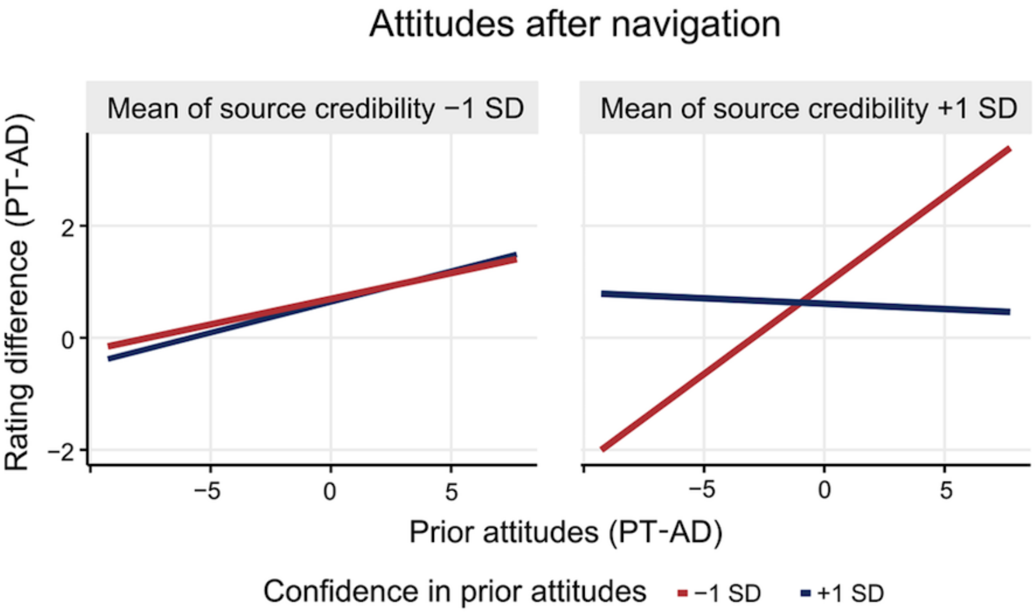
[f] PT: psychotherapy.

[g] PA: prior attitudes.

[h] Not applicable.

We expected that high (vs low) confidence would lead to higher confirmation bias and decreased attitude change; therefore, for people who hold their attitudes with high (vs low) confidence, prior attitudes should be more strongly associated with attitudes after navigation (H2d). The expected interaction between confidence and prior attitudes was not significant (Step 2 in Table 4). However, the association between confidence and prior attitudes depended on source credibility (Step 3 in Table 4). To disentangle this 3-way interaction, simple slopes were tested on low (−1 SD) and high (+1 SD) levels of source credibility ratings and confidence ratings. This revealed a strong association between prior attitudes and treatment efficacy ratings after navigation for participants with lower ratings of confidence (−1 SD) and high source credibility ratings (+1 SD; beta=.34, SE 0.13; *P*=.01) but no association for high confidence ratings (+1 SD) and low source credibility ratings (−1 SD; beta=.11, SE 0.06; *P*=.053). There was also no association with low confidence (−1 SD) and low source credibility ratings (−1 SD; beta=.08, SE 0.10; *P*=.42) and with high confidence (+1 SD) and high source credibility ratings (+1 SD; beta=−.04, SE 0.06; *P*=.50; Figure 9).

Figure 9. Prior attitudes, confidence and source credibility, and treatment efficacy ratings after navigation, with negative values on all axes indicating a preference for antidepressants over psychotherapy. AD: antidepressants; PT: psychotherapy.



Tag Popularity of Treatments in the Social Tag Cloud

Tag Selection

In H3a, we expected that popular treatment tags would be selected more often, independent of prior attitudes. To test this, we used a logistic regression model as described in the previous confidence section on tag selection (see Table 2, Step 1). Tag popularity was the only significant predictor for the proportion of selected antidepressant tags (beta=.44, SE 0.11; *P*<.001). This supports H3a, as participants selected a larger proportion of popular tags in the tag cloud. They did this independent of their prior attitudes, as prior attitudes were not associated with tag selection.

Blog Post Selection

We also expected that participants would select more blog posts when related tags were more popular (H3b). We tested this with the logistic regression as described in the previous confidence section on blog post selection (see Table 3). This showed that participants selected a larger proportion of blog posts when related tags were popular in the tag cloud, supporting H3b (beta=.44, SE 0.11; *P*<.001; Table 3, Step 1).

Attitude Change

We expected in H3c that the attitude change would depend on tag popularity. More popular tags for a treatment should be associated with higher ratings of treatment efficacy. We conducted a linear regression analysis, as described in the previous confidence section (also see Table 4). We did not find an influence of tag popularity on efficacy ratings after navigation (beta=−.05, SE 0.12; *P*=.64; see Table 4, Step 1).

Source Credibility of the Tagging Community

Tag Selection

We expected that when the tagging community comprises experts (vs novices), participants click on more tags (H4b). To test this, we conducted a negative binomial regression with the continuous, centered predictors (source credibility score, prior attitudes, and confidence score), the dichotomous predictor (tag popularity; 0=psychotherapy tags popular and 1=antidepressant tags popular), and the dependent variable (number of selected tags; Table 5). We did not find support for H4b, as the total number of selected tags was not associated with source credibility ratings.

Table 5. Tags and blog posts selected.

Predictor	Tags selected[a]			Blog posts selected[b]		
	Beta[c]	SE	*P* value	Beta	SE	*P* value
Intercept	1.81	0.07	<.001	1.16	0.47	.01
PA[d]	.01	0.03	.63	.03	0.04	.48
Confidence score	−.02	0.05	.62	.06	0.08	.43
Tag popularity	.14	0.10	.18	−.34	0.18	.06
Source credibility score	.03	0.04	.51	.02	0.07	.79

[a] Model fit: χ^2_4=2.7; *P*=.61.

[b] Model fit: χ^2_6=5.1; *P*=.28.

[c] Continuous predictors were centered.

[d] PA: prior attitudes.

Blog Post Selection

We expected that when the tagging community comprises experts (vs novices), participants click on more blog posts (H4c). We conducted another negative binomial regression with the continuous, centered predictors (source credibility score, prior attitudes, confidence score), the dichotomous predictor (tag popularity: 0=psychotherapy tags popular and 1=antidepressant tags popular), and the dependent variable (total number of selected blog posts; Table 5). As with the number of selected tags, perceived source credibility did not predict the total number of selected blog posts, not supporting H4c.

Attitude Change

When the tagging community comprised experts (vs novices), we expected that participants should change their prior attitudes to a greater degree (H4d). We conducted a linear regression analysis with efficacy ratings before navigation (PT−AD) as a covariate [52] and included the predictors tag popularity (0=psychotherapy tags popular and 1=antidepressant tags popular), prior attitudes, confidence, and source credibility ratings. As a criterion, we included the efficacy rating difference (PT−AD) after navigation. The covariate and the continuous predictor variables were centered [53]. High perceived source credibility was associated with higher treatment efficacy ratings after navigation (Table 6), supporting H4d.

Table 6. Treatment efficacy ratings (antidepressants+psychotherapy) after navigation.

Predictor	Ratings T2 (AD[a]+PT[b])[c]		
	Beta[d]	SE	*P* value
Intercept	10.26	0.12	<.001
Efficacy ratings before navigation (AD+PT)	.76	0.05	<.001
Tag popularity	.02	0.18	.93
PA[e]	.08	0.04	.06
Confidence in PA score	.14	0.09	.12
Source credibility score	.24	0.07	<.001

[a] AD: antidepressants.

[b] PT: psychotherapy.

[c] Model fit: adjusted R^2=.16, $F_{5,220}$=9.47; *P*<.001.

[d] Continuous predictors and criteria were centered.

[e] PA: prior attitudes.

Discussion

Principal Findings

With this randomized, controlled study, we aimed to investigate prior attitudes about antidepressants and psychotherapy and the tendency to confirm prior attitudes when selecting and evaluating mental health–related information. We presented 3 factors to counter confirmation bias: popularity of treatment tags in a social tag cloud, confidence, and the source credibility of the tagging community. We expected that people would select and favorably evaluate attitude-inconsistent content when confidence was low (vs high). In addition, we expected that *source credibility* and *tag popularity* should influence selection of tags independent of prior attitudes. We could not replicate the confidence manipulation (Study 3 in [17]) and participants did not distinguish source credibility as presented by banners; therefore, we used manipulation check scores for correlational analyses.

As expected, people in the German population rated psychotherapy as more effective than antidepressants, and they reported according beliefs. Increasing tag popularity increased selection of tags, independent of prior attitudes and confidence. In contrast to our expectations, higher source credibility was not associated with increased tag or blog post selection. Participants with high confidence were more open to select attitude-inconsistent blog posts, which is in line with the defense motivation account but not with the accuracy motivation account we had expected [7]. Moreover, we found that people with low confidence rated treatment efficacy in accordance with their prior attitudes but only when perceived source credibility was high.

Social Tags to Reduce Confirmation Bias

We expected that social tag clouds are a nonintrusive interface to circumvent prior attitudes, and popular tags would be selected more often independent of prior attitudes. We found that people selected popular tags and related blog posts more often. We think that these findings highlight the important role of popular content on the Web and also in the context of mental health–related selection of information. When two treatment options are presented to a searcher, searchers will be guided by more popular information, even independent of their prior attitudes. This could help to design Web-based platforms in which it is desirable to minimize the influence of prior attitudes and maximize the influence of a community.

A thorough discussion about nudges is beyond the scope of this paper, but we consider implications of implementing tag clouds as nudges. Though nudges are controversial in general [54], educational nudges aiming to aid people in making better decisions are less controversial [54]. Moreover, in the health context, it is argued that it is impossible not to be influenced by policies of different stakeholders in general [55]. The way in which tools such as tag clouds influence behavior might be considered more controversial as large tags automatically attract the searchers' attention [25], thus influencing information selection [21], and therefore tags may restrict deliberate individual agency [54,56].

Defense Motivation in Mental Health–Related Information Search

We expected that people would be guided by accuracy motivation when searching for mental health–related information. People would strive to select and evaluate information that is objectively correct, regardless of their prior attitudes. In contrast to this, the pattern of results suggests that information searchers were defense motivated, and they tended to confirm their prior attitudes to avoid dissonant cognitions and to maintain a positive view of themselves [7,10,57].

This was reflected in blog post selection and resulting attitude change. We found that low confidence was associated with selecting attitude-consistent blog posts, which suggests that participants may have felt increased threat under low confidence.

The findings on attitude change provide further support for the defense motivation account. People with high confidence were expected to change their attitudes in line with their prior attitudes. However, we found the opposite. When confidence was low, not high, people's attitudes after navigation were polarized in line with their prior attitudes. However, in contrast to blog post selection, this pattern was only found when source credibility was high but not when source credibility was low. This suggests that attitude-inconsistent information could have posed a double threat when source credibility was high, in combination with low confidence. In all other instances, there was no association between prior attitudes and attitude change.

What follows from defensive processing? Not only when information presents a direct threat (eg, antismoking images) but also when different treatment options are available, prior attitudes have an impact on Web-based information search. When information acknowledges prior attitudes of the reader, the need to maintain a positive self-view can be reduced, and the reader becomes more open to attitude-inconsistent information [58,59]. Therefore, content authors could anticipate the attitudes of their readers when providing health information and acknowledge existing attitudes and views before providing potentially conflicting information.

Source Credibility and Confirmation Bias

People do sometimes recognize source credibility on the Web [33,34,36]; however, participants did not rate practitioners with years of experience as more credible compared with students of health-related subjects in their first semester. One possible explanation for this is that the banners on top of the page were too subtle.

Moreover, for student samples (vs nonstudent samples) [35] and content that is presented on common websites (vs user-generated content) [34], searchers perceive experts as more credible. This might explain that for this representative sample on a specific tagging platform, people did not distinguish high from low source credibility.

For content authors, this finding underlines the importance to consider the target audience as well as the impact of the type of platform that is being used to convey health-related messages. Although information searchers with high educational background or searchers on general websites respond more to expertise, searchers on sites presenting user-generated content (eg, forums or blog posts) respond more to demographic similarity to the searcher [34], and nonstudent searchers respond less to expertise when judging source credibility.

Attitude Confidence and Confirmation Bias

A recent study showed that individuals with depressive and anxiety symptoms exhibited lower confidence in a decision-making task [15]. In this study, people with lower confidence evaluated information content in line with their prior attitudes, when the source of information was highly credible. Therefore, when searchers particularly perceive information as highly credible, individuals with depressive or anxiety symptoms might be prone to select attitude-consistent information. This should also be tested by future studies.

Public Attitudes Toward Antidepressants and Psychotherapy

As for student [20] and representative samples in Germany [38], we also expected prior attitudes to be more positive for psychotherapy than for antidepressants, and we found an according moderate effect. The results about the specific beliefs show that people are not satisfied with the current accessibility of mental health care services, and the German population seems to have specific beliefs when it comes to side effects of antidepressants. However, side effects that can be found in the literature, such as nausea, insomnia, somnolence, fatigue, sexual dysfunction, and weight gain [60,61], were rarely associated with antidepressants.

Limitations

According to the Federal Office of Statistics, the sample from this study is representative for gender and age, but participants with lower education, such as people with a qualified job, are underrepresented, whereas participants with a university degree are slightly overrepresented [62]. Therefore, the results of this study should be interpreted with caution for people with lower-level education. The recruitment process of the panel company uses Web-based campaigns, search engine marketing, and offline recruitment, where participants register at a portal through which they can enroll for studies that match their demographics. Therefore, it should be noted that this sample is restricted to Web-based users of the German population.

This study suggests that the results for confidence and its interplay with source credibility are in line with predictions of defense motivation; however, because of the correlational design, potential correlated confounding influences could be at work and could potentially have been overlooked.

Moreover, all blog posts highlighted the efficacy aspect of prior attitudes, whereas other important issues such as side effects or treatment of psychological causes were not mentioned in the blog posts. Thus, only one aspect related to prior attitudes, namely treatment efficacy, was addressed in the blog posts. In addition, all blog posts were formulated positively, such that information revealing limitations and boundary conditions of the treatments were addressed in the blog posts.

As age could be an important covariate in this study, we exploratively checked the influence of age for each dependent variable; however, age was not a significant predictor in none of the analyses.

Conclusions

We presented correlational support for the defense motivation account in health-related search. That is, participants tended to confirm their prior attitudes when searching for information. We presented factors that influence this confirmation bias. First, social tags reduced the influence of prior attitudes, and second, attitude confidence increased confirmation bias when source credibility was high. These findings have many implications for content creators, who should acknowledge existing attitudes in persuasive communication and consider demographics of their audience as well as the type of platform where content is published. Future studies should test whether this result extends to other health-related domains, beyond treatment of depression, and to other information platforms as well. Furthermore, it would be highly interesting to compare treatment attitudes toward internet-based psychotherapy including different delivery modes.

Acknowledgments

This study was funded by the Knowledge Media Research Center. The authors thank André Klemke and Manfred Knobloch who designed the tagging platform.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the social tagging environment used in the current study.

[PDF File (Adobe PDF File), 511KB - jmir_v21i4e11081_app1.pdf]

Multimedia Appendix 2

Original data and R analysis file.

[ZIP File (Zip Archive), 924KB - jmir_v21i4e11081_app2.zip]

Multimedia Appendix 3

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 875KB - jmir_v21i4e11081_app3.pdf]

References

1. Liao QV, Fu WT. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. New York, NY: ACM; 2013 Presented at: CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 27-May 02, 2013; Paris, France p. 2359-2368.
2. Kimmerle J, Bientzle M, Cress U, Flemming D, Greving H, Grapendorf J, et al. Motivated processing of health-related information in online environments. In: Buder J, Hesse F, editors. Informational Environments. Cham, Switzerland: Springer; 2017:75-96.
3. Del Vicario M, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W. Modeling confirmation bias and polarization. Sci Rep 2017 Dec 11;7:40391 [FREE Full text] [doi: 10.1038/srep40391] [Medline: 28074874]
4. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 1998 Nov;2(2):175-220. [doi: 10.1037/1089-2680.2.2.175]
5. White RW, Horvitz E. Belief dynamics and biases in web search. ACM Trans Inf Syst 2015 May 15;33(4):1-46. [doi: 10.1145/2746229]
6. Klayman J. Varieties of confirmation bias. Psychol Learn Motiv 1995;32:385-418 [FREE Full text] [doi: 10.1016/S0079-7421(08)60315-1]
7. Hart W, Albarracín D, Eagly AH, Brechan I, Lindberg MJ, Merrill L. Feeling validated versus being correct: a meta-analysis of selective exposure to information. Psychol Bull 2009 Jul;135(4):555-588 [FREE Full text] [doi: 10.1037/a0015701] [Medline: 19586162]
8. Wood W. Attitude change: persuasion and social influence. Annu Rev Psychol 2000;51:539-570. [doi: 10.1146/annurev.psych.51.1.539] [Medline: 10751980]
9. Leary MR, Terry ML, Batts Allen A, Tate EB. The concept of ego threat in social and personality psychology: is ego threat a viable scientific construct? Pers Soc Psychol Rev 2009 Aug;13(3):151-164. [doi: 10.1177/1088868309342595] [Medline: 19648508]
10. Jonas E, Schulz-Hardt S, Frey D. Giving advice or making decisions in someone else's place: the influence of impression, defense, and accuracy motivation on the search for new information. Pers Soc Psychol Bull 2005 Jul 9;31(7):977-990. [doi: 10.1177/0146167204274095] [Medline: 15951368]
11. Moore DA, Healy PJ. The trouble with overconfidence. Psychol Rev 2008 Apr;115(2):502-517. [doi: 10.1037/0033-295X.115.2.502] [Medline: 18426301]
12. Koriat A, Lichtenstein S, Fischhoff B. Reasons for confidence. J Exp Psychol Hum Learn Mem 1980;6(2):107-118. [doi: 10.1037/0278-7393.6.2.107]
13. Zell E, Krizan Z. Do people have insight into their abilities? A metasynthesis. Perspect Psychol Sci 2014 Mar;9(2):111-125. [doi: 10.1177/1745691613518075] [Medline: 26173249]
14. Park JH, Konana P, Gu B, Kumar A, Raghunathan R. Confirmation bias, overconfidence, and investment performance: evidence from stock message boards. oc Sci Res Netw Electron Pap Collect 2010 Jul 12:59. [doi: 10.2139/ssrn.1639470]
15. Rouault M, Seow T, Gillan CM, Fleming SM. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biol Psychiatry 2018 Sep 15;84(6):443-451 [FREE Full text] [doi: 10.1016/j.biopsych.2017.12.017] [Medline: 29458997]
16. Johnson DR, Murphy MP, Messer RM. Reflecting on explanatory ability: a mechanism for detecting gaps in causal knowledge. J Exp Psychol Gen 2016 May;145(5):573-588. [doi: 10.1037/xge0000161] [Medline: 26999047]
17. Petty RE, Briñol P, Tormala ZL. Thought confidence as a determinant of persuasion: the self-validation hypothesis. J Pers Soc Psychol 2002 May;82(5):722-741. [Medline: 12003473]
18. Briñol P, Petty RE. Persuasion: insights from the self-validation hypothesis. In: Zanna MP, editor. Advances in Experimental Social Psychology. San Diego, CA, US: Elsevier Academic Press; 2009:69-118.
19. Muchnik L, Aral S, Taylor SJ. Social influence bias: a randomized experiment. Science 2013 Aug 9;341(6146):647-651. [doi: 10.1126/science.1240466] [Medline: 23929980]
20. Schweiger S, Oeberst A, Cress U. Confirmation bias in web-based search: a randomized online study on the effects of expert information and social tags on information search and evaluation. J Med Internet Res 2014 Mar 26;16(3):e94 [FREE Full text] [doi: 10.2196/jmir.3044] [Medline: 24670677]
21. Cress U, Held C, Kimmerle J. The collective knowledge of social tags: direct and indirect influences on navigation, learning, and information processing. Comput Educ 2013 Jan;60(1):59-73. [doi: 10.1016/j.compedu.2012.06.015]
22. Seitlinger P, Ley T, Albert D. An Implicit-Semantic Tag Recommendation Mechanism for Socio-Semantic Learning Systems. Berlin, Heidelberg: Springer; 2012 Presented at: IFIP WG 3.4 International Conference on Open and Social Technologies for Networked Learning; July 30-August 3, 2012; Tallinn, Estonia p. 41-46. [doi: 10.1007/978-3-642-37285-8_5]
23. Ley T, Seitlinger P. Dynamics of human categorization in a collaborative tagging system: how social processes of semantic stabilization shape individual sensemaking. Comput Human Behav 2015 Oct;51(Pt A):140-151 [FREE Full text] [doi: 10.1016/j.chb.2015.04.053] [Medline: 26566299]
24. Seitlinger P, Ley T, Albert D. Verbatim and Semantic Imitation in Indexing Resources on the Web: A Fuzzy-trace Account of Social Tagging. Appl. Cognit. Psychol 2014 Aug 29;29(1):32-48. [doi: 10.1002/acp.3067]

25. Bateman S, Gutwin C, Nacenta M. Seeing things in the clouds: the effect of visual features on tag cloud selections. 2008 Presented at: Proceedings of the nineteenth ACM conference on Hypertext and Hypermedia; June 19-21, 2008; Pittsburgh, PA, USA p. 193-202. [doi: 10.1145/1379092.1379130]
26. Gligorov R, Hildebrand M, van Ossenbruggen J, Schreiber G, Aroyo L. On the role of user-generated metadata in audio visual collections. 2011 Presented at: Proceedings of the sixth international conference on Knowledge capture; June 26-29, 2011; Banff, Alberta, Canada p. 145-152. [doi: 10.1145/1999676.1999702]
27. Lohmann S, Ziegler J, Tetzlaff L. Comparison of tag cloud layouts: task-related performance and visual exploration. In: Gross T, editor. Human-Computer Interaction – INTERACT 2009. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2009:392-404.
28. Kuo BY, Hentrich T, Good BM, Wilkinson MD. Tag clouds for summarizing web search results. New York, NY: ACM Press; 2007 Presented at: Proceedings of the 16th international conference on World Wide Web; May 08-12, 2007; Banff, Alberta, Canada p. 1203-1204. [doi: 10.1145/1242572.1242766]
29. Cress U, Held C. Harnessing collective knowledge inherent in tag clouds. J Comput Assist Learn 2012 Aug 23;29(3):235-247. [doi: 10.1111/j.1365-2729.2012.00491.x]
30. Held C, Kimmerle J, Cress U. Learning by foraging: the impact of individual knowledge and social tags on web navigation processes. Comput Human Behav 2012 Jan;28(1):34-40. [doi: 10.1016/j.chb.2011.08.008]
31. Bohner G, Dykema-Engblade AD, Tindale RS, Meisenhelder H. Framing of Majority and Minority Source Information in Persuasion. Social Psychology 2008 Jan;39(2):108-116. [doi: 10.1027/1864-9335.39.2.108]
32. Darke PR, Chaiken S, Bohner G, Einwiller S, Erb HP, Hazlewood JD. Accuracy Motivation, Consensus Information, and the Law of Large Numbers: Effects on Attitude Judgment in the Absence of Argumentation. Pers Soc Psychol Bull 2016 Jul 02;24(11):1205-1215. [doi: 10.1177/01461672982411007]
33. Kammerer Y, Gerjets P. The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. Int J Human-Comput Interact 2014 Jan 31;30(3):177-191. [doi: 10.1080/10447318.2013.846790]
34. Ma TJ, Atkin D. User generated content and credibility evaluation of online health information: a meta analytic study. Telemat Inform 2017 Aug;34(5):472-486. [doi: 10.1016/j.tele.2016.09.009]
35. Yang Q, Beatty M. A meta-analytic review of health information credibility:belief in physicians or belief in peers? Health Inf Manag 2016 Aug;45(2):80-89. [doi: 10.1177/1833358316639432] [Medline: 27105476]
36. Kammerer Y, Bråten I, Gerjets P, Strømsø HI. The role of Internet-specific epistemic beliefs in laypersons' source evaluations and decisions during Web search on a medical issue. Comput in Human Behav 2013 May;29(3):1193-1203. [doi: 10.1016/j.chb.2012.10.012]
37. O'Grady L, Wathen CN, Charnaw-Burger J, Betel L, Shachak A, Luke R, et al. The use of tags and tag clouds to discern credible content in online health message forums. Int J Med Inform 2012 Jan;81(1):36-44. [doi: 10.1016/j.ijmedinf.2011.10.001] [Medline: 22030035]
38. Angermeyer MC, Matschinger H, Schomerus G. Attitudes towards psychiatric treatment and people with mental illness: changes over two decades. Br J Psychiatry 2013 Aug;203(2):146-151 [FREE Full text] [doi: 10.1192/bjp.bp.112.122978] [Medline: 23787060]
39. Schomerus G, Schwahn C, Holzinger A, Corrigan PW, Grabe HJ, Carta MG, et al. Evolution of public attitudes about mental illness: a systematic review and meta-analysis. Acta Psychiatr Scand 2012 Jun;125(6):440-452. [doi: 10.1111/j.1600-0447.2012.01826.x] [Medline: 22242976]
40. Angermeyer MC, Breier P, Dietrich S, Kenzine D, Matschinger H. Public attitudes toward psychiatric treatment. An international comparison. Soc Psychiatry Psychiatr Epidemiol 2005 Nov;40(11):855-864. [doi: 10.1007/s00127-005-0958-x] [Medline: 16215879]
41. Jorm AF, Nakane Y, Christensen H, Yoshioka K, Griffiths KN, Wata Y. Public beliefs about treatment and outcome of mental disorders: a comparison of Australia and Japan. BMC Med 2005 Jul 9;3:12 [FREE Full text] [doi: 10.1186/1741-7015-3-12] [Medline: 16004615]
42. Jorm A. Mental health literacy. Public knowledge and beliefs about mental disorders. Br J Psychiatry 2000 Nov;177:396-401. [Medline: 11059991]
43. Jorm AF, Christensen H, Griffiths KM. The public's ability to recognize mental disorders and their beliefs about treatment: changes in Australia over 8 years. Aust N Z J Psychiatry 2006 Jan;40(1):36-41. [doi: 10.1080/j.1440-1614.2006.01738.x] [Medline: 16403035]
44. van Schaik DJ, Klijn A, van Hout HP, van Marwijk HW, Beekman A, de Haan M, et al. Patients' preferences in the treatment of depressive disorder in primary care. Gen Hosp Psychiatry 2004;26(3):184-189. [doi: 10.1016/j.genhosppsych.2003.12.001] [Medline: 15121346]
45. Van der Auwera S, Schomerus G, Baumeister S, Matschinger H, Angermeyer M. Approval of psychotherapy and medication for the treatment of mental disorders over the lifespan. An age period cohort analysis. Epidemiol Psychiatr Sci 2017 Dec;26(1):61-69. [doi: 10.1017/S2045796015001134] [Medline: 26753632]

46. Angermeyer MC, van der Auwera S, Carta MG, Schomerus G. Public attitudes towards psychiatry and psychiatric treatment at the beginning of the 21st century: a systematic review and meta-analysis of population surveys. World Psychiatry 2017 Feb;16(1):50-61 [FREE Full text] [doi: 10.1002/wps.20383] [Medline: 28127931]
47. De Maat S, Dekker J, Schoevers R, De Jonghe F. Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Psychother Res 2006 Oct;16(5):566-578. [doi: 10.1080/10503300600756402]
48. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
49. Amick HR, Gartlehner G, Gaynes BN, Forneris C, Asher GN, Morgan LC, et al. Comparative benefits and harms of second generation antidepressants and cognitive behavioral therapies in initial treatment of major depressive disorder: systematic review and meta-analysis. Br Med J 2015 Dec 8;351:h6019 [FREE Full text] [doi: 10.1136/bmj.h6019] [Medline: 26645251]
50. Menard S. Applied Logistic Regression Analysis. California: Sage; Oct 9, 2001:128.
51. Peng CY, Lee KL, Ingersoll GM. An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research 2010 Apr 02;96(1):3-14. [doi: 10.1080/00220670209598786]
52. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
53. Aiken LS, West SG, Reno RR. Multiple Regression: Testing and Interpreting Interactions. California: Sage; Jan 1, 1991:224.
54. Sunstein CR. The ethics of nudging. SSRN J 2014 Nov 20:1-50. [doi: 10.2139/ssrn.2526341]
55. Quigley M. Nudging for health: on public policy and designing choice architecture. Med Law Rev 2013;21(4):588-621 [FREE Full text] [doi: 10.1093/medlaw/fwt022] [Medline: 24081425]
56. Le Clainche C, Tubeuf S. Nudging, intervening or rewarding: a discussion on the constraints and the degree of control on health status. Polit Philos Econ 2015 Sep 9;15(2):170-189 [FREE Full text] [doi: 10.1177/1470594X15600832]
57. Festinger L. Conflict, decision, and dissonance. Palo Alto, CA: Stanford University Press; 1964:163.
58. Harris PR, Napper L. Self-affirmation and the biased processing of threatening health-risk information. Pers Soc Psychol Bull 2005 Sep;31(9):1250-1263. [doi: 10.1177/0146167205274694] [Medline: 16055644]
59. van Koningsbruggen GM, Das E, Roskos-Ewoldsen DR. How self-affirmation reduces defensive processing of threatening health information: evidence at the implicit level. Health Psychol 2009 Sep;28(5):563-568. [doi: 10.1037/a0015610] [Medline: 19751082]
60. Papakostas GI. Tolerability of modern antidepressants. J Clin Psychiatry 2008;69(Suppl E1):8-13 [FREE Full text] [Medline: 18494538]
61. Khawam EA, Laurencic G, Malone Jr DA. Side effects of antidepressants: an overview. Cleve Clin J Med 2006 Apr;73(4):351-356. [Medline: 16610395]
62. Jahrbuch Deutschland. In: Statistisches Bundesamt. Wiesbaden, Hessen: Statistisches Bundesamt; 2013.

Edited by G Eysenbach; submitted 18.05.18; peer-reviewed by M Noll-Hussong, J Apolinário-Hagen, S Brigitte; comments to author 24.09.18; revised version received 03.12.18; accepted 10.12.18; published 23.04.19.

Please cite as:
Schweiger S, Cress U
How Confidence in Prior Attitudes, Social Tag Popularity, and Source Credibility Shape Confirmation Bias Toward Antidepressants and Psychotherapy in a Representative German Sample: Randomized Controlled Web-Based Study
J Med Internet Res 2019;21(4):e11081
URL: http://www.jmir.org/2019/4/e11081/
doi:10.2196/11081
PMID:31012865

Original Paper

Symptoms Prompting Interest in Celiac Disease and the Gluten-Free Diet: Analysis of Internet Search Term Data

Benjamin Lebwohl[1], MD, MS; Elad Yom-Tov[2,3], PhD

[1]Celiac Disease Center, Columbia University, New York, NY, United States
[2]Microsoft Research, Herzeliya, Israel
[3]Technion, Haifa, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
13 Shenkar Street
Herzeliya, 46733
Israel
Phone: 972 747111359
Email: eladyt@yahoo.com

Abstract

Background: Celiac disease, a common immune-based disease triggered by gluten, has diverse clinical manifestations, and the relative distribution of symptoms leading to diagnosis has not been well characterized in the population.

Objective: This study aimed to use search engine data to identify a set of symptoms and conditions that would identify individuals at elevated likelihood of a subsequent celiac disease diagnosis. We also measured the relative prominence of these search terms before versus after a search related to celiac disease.

Methods: We extracted English-language queries submitted to the Bing search engine in the United States and identified those who submitted a new celiac-related query during a 1-month period, without any celiac-related queries in the preceding 9 months. We compared the ratio between the number of times that each symptom or condition was asked in the 14 days preceding the first celiac-related query of each person and the number of searches for that same symptom or condition in the 14 days after the celiac-related query.

Results: We identified 90,142 users who made a celiac-related query, of whom 6528 (7%) exhibited sustained interest, defined as making a query on more than 1 day. Though a variety of symptoms and associated conditions were also queried before a celiac-related query, the maximum area under the receiver operating characteristic curve was 0.53. The symptom most likely to be queried more before than after a celiac-related query was diarrhea (query ratio [QR] 1.28). Extraintestinal symptoms queried before a celiac disease query included headache (QR 1.26), anxiety (QR 1.10), depression (QR 1.03), and attention-deficit hyperactivity disorder (QR 1.64).

Conclusions: We found an increase in antecedent searches for symptoms known to be associated with celiac disease, a rise in searches for depression and anxiety, and an increase in symptoms that are associated with celiac disease but may not be reported to health care providers. The protean clinical manifestations of celiac disease are reflected in the diffuse nature of antecedent internet queries of those interested in celiac disease, underscoring the challenge of effective case-finding strategies.

(J Med Internet Res 2019;21(4):e13082) doi:10.2196/13082

KEYWORDS

celiac disease; gluten; epidemiology

Introduction

Celiac disease is a multisystem immune-based enteropathy characterized by autoantibodies to tissue transglutaminase and villous atrophy that is triggered by the ingestion of gluten in genetically predisposed individuals [1]. Present in nearly 1% of the US population, the seroprevalence of celiac disease has risen markedly in recent decades [2]. However, this rise in celiac disease prevalence has been eclipsed by a more dramatic rise in the avoidance of gluten among individuals who do not have celiac disease or have not been tested for celiac disease [3]. The reasons for gluten avoidance are manifold and include gastrointestinal symptoms [4], the avoidance of cardiometabolic complications [5], cognitive health [6], and treatment of autoimmune disease [7].

Although classical celiac disease consists of a malabsorption phenotype characterized by diarrhea and weight loss, the majority of patients with celiac disease are diagnosed with a *nonclassical* presentation that includes a heterogeneous set of symptoms and signs including osteoporosis, anemia, abnormal liver enzymes, neuropathy, infertility, and others [8]. Given these diverse clinical manifestations and the lack of signs and symptoms that are sensitive and specific for celiac disease, diagnosis can be elusive, and there is frequently a long delay between the onset of symptoms and the diagnosis of celiac disease. In 1 survey of adults in the United States, patients with celiac disease had symptoms for a mean of 11 years before diagnosis [9]. Patients with extraintestinal symptoms tend to have a longer diagnostic delay compared with those with intestinal symptoms [10].

Aside from clinical examination, researchers have used electronic medical records to screen for celiac disease, showing that analysis of free text included therein could aid in the early discovery of celiac disease [11]. Recently, another novel method for the ascertainment of symptoms preceding diagnoses has been proposed through the analysis of search engine queries [12]. For instance, search engine query analysis has been used to identify symptoms of pancreatic adenocarcinoma months before the disease diagnosis [13]. In this study, we examined search query data with the aim of identifying symptoms and conditions that are associated with a subsequent query for celiac disease and celiac disease–specific searches. We hypothesized that searches related to the modes of presentation of celiac disease would precede searches for celiac disease and/or the gluten-free diet. We aimed to measure the relative prominence of these search terms before versus after a search related to celiac disease. We also aimed to identify a set of symptoms and conditions that would identify individuals at elevated likelihood of a subsequent celiac disease diagnosis.

Methods

We extracted all English-language queries submitted to the Bing search engine between January 1, 2017, and October 31, 2017, by people in the United States. For each query, we extracted the time and date of the query, its text, an anonymous user identifier, and the zip code of the asker. Bing data are estimated to be a representative sample of US internet users [14].

Celiac-related queries (CRQs) were those queries that contained the words “celiac” or “gluten.” The queries were filtered to include only those queries by users who were active since at least September 1, 2017, and used CRQs during the month of October 2017, but not in the previous 9 months.

In addition, we identified queries that could indicate celiac disease by finding those queries that contained 1 or more of the following terms: marsh score, duodenal biopsy, intestinal biopsy, beyond celiac, celiac disease foundation, tissue transglutaminase (also as an acronym: TTG), gliadin antibody, celiac clinical trials, celiac trials, or gluten trials.

Some people identified themselves in their queries as having celiac disease, through queries such as “I have celiac disease, can I eat rice?”. To identify these self-identified users [15], we found all mentions of “I have celiac” or “I was diagnosed with celiac” and manually inspected each to exclude irrelevant queries (eg, “do I have celiac?”). Obviously, not all people who have celiac identify themselves in their queries, but this subset of the population is likely composed of celiac patients, and we calculated the prevalence of CRQs in this subset.

We defined people with a passing interest in celiac disease as those who made CRQs during only 1 day. This contrasts with people with a sustained interest who made CRQs over more than 1 day.

Symptoms mentioned in queries were identified by matching the text of queries to a list of 195 symptoms and their synonyms, as developed by Yom-Tov and Gabrilovich [16]. Similarly, medical conditions were found by extracting all 5521 diseases and their synonyms that appear in Wikipedia [17].

Recipe searches provide a representative sample of the dietary consumption of individuals [18]. Therefore, to evaluate the changes in diet made by people with CRQs, we followed the methodology used previously [18] to identify queries for recipes and map them to ingredients therein.

We attempted to identify users with sustained interest from all users using their queries. To do this, we represented each user by the number of times they queried for each symptom and each medical condition before the first CRQ. A predictive model using either linear regression or random forest with 50 trees was constructed and tested using 10-fold cross-validation [19].

This study was approved by the Behavioral Sciences Research Ethics Committee of the Technion, approval number 2018-032.

Results

Of 90,142 users with at least 1 CRQ, 83,614 users (93%) were found with passing interest and 6528 (7%) exhibited sustained interest. Of the 6528 people who had a sustained interest in celiac disease, 104 (1.6%) entered at least one celiac indicator (see Table 1) compared with 336 (0.7%) who had a passing interest and 0.001% in the general population of Bing users.

Table 1. Symptoms and conditions that appear with the highest probability in the 14 days before the first celiac-related query compared with the 14 days after it in the entire population and in the sustained user population.

Category	All celiac queries		Sustained celiac queries	
	Item	Ratio	Item	Ratio
Symptoms	Steatorrhea	1.88	Diarrhea	1.34
	Dyspepsia	1.76	Cough	1.28
	Exophthalmos	1.63	Headache	1.27
	Vomiting	1.62	Bloating	1.26
	Stomach ache	1.57	Anxiety	1.10
	Diarrhea	1.50	Bleeding	1.08
	Xerostomia	1.49	Weight loss	1.03
	Flatulence	1.48	Depression	1.03
	Abdominal pain	1.46	Pain	1.01
	Bloating	1.45	Itch	1.00
Condition	Lactose intolerance	3.05	Autoimmunity	3.21
	Inflammatory bowel disease	2.58	Attention-deficit hyperactivity disorder	1.64
	Malabsorption	2.32	Hypothyroidism	1.38
	Peptic ulcer	2.22	Gastroesophageal reflux disease	1.33
	Irritable bowel syndrome	2.19	Asthma	1.30
	Food intolerance	2.10	Influenza	1.24
	Crohn disease	2.02	Migraine	1.22
	Digestive disease	1.96	Colitis	1.21
	Polycystic ovary syndrome	1.95	Systemic lupus erythematosus	1.13
	Peritonitis	1.92	Alzheimer disease	1.13

Only 31 users identified themselves as having celiac disease based on a self-identified query (eg, “I have celiac”). Among people with a sustained interest in celiac disease, 0.12% (8/6528) were those who identified themselves as having celiac disease compared with 0.03% (31/90142) in the general population (a ratio of 3.5). Thus, similar to the findings of Ofran et al [20], a sustained interest in celiac disease can be considered a proxy for having the condition or for a caregiver of a patient.

We attempted to identify users with either a passing or sustained interest in celiac disease, comparing them with all other users based on antecedent queries. In both cases, the area under the receiver operating characteristic curve was 0.53 or less, indicating that we could not distinguish the 2 classes based on symptoms and conditions searched.

To investigate the symptoms most associated with initiation of search for CRQs, we compared the ratio between the number of times that each symptom or condition were asked in the 14 days preceding the first CRQ of each person and the number of times they were mentioned from 14 days before the first CRQ until 14 days after it. Table 1 shows the symptoms and conditions with the highest before-to-after ratio.

Figure 1 shows the fraction of queries for “diarrhea,” the top-ranked symptom among sustained users, over time compared with the fraction of all queries. As the figure shows, interest in this symptom begins to rise only approximately 2 weeks before the first CRQ and rises dramatically in the few days before it.

As noted above, we identified queries for recipes and compared these recipes from 14 days before the first CRQ with recipes from 14 days after it. Table 2 shows the recipes and ingredients that increased in searches and those that decreased.

Figure 1. Fraction of queries for diarrhea over time, compared with the fraction of all queries. Day zero is the first celiac-related query. A ratio greater than 1 indicates that the queries for diarrhea are more common than could be expected. CRQ: celiac-related query.

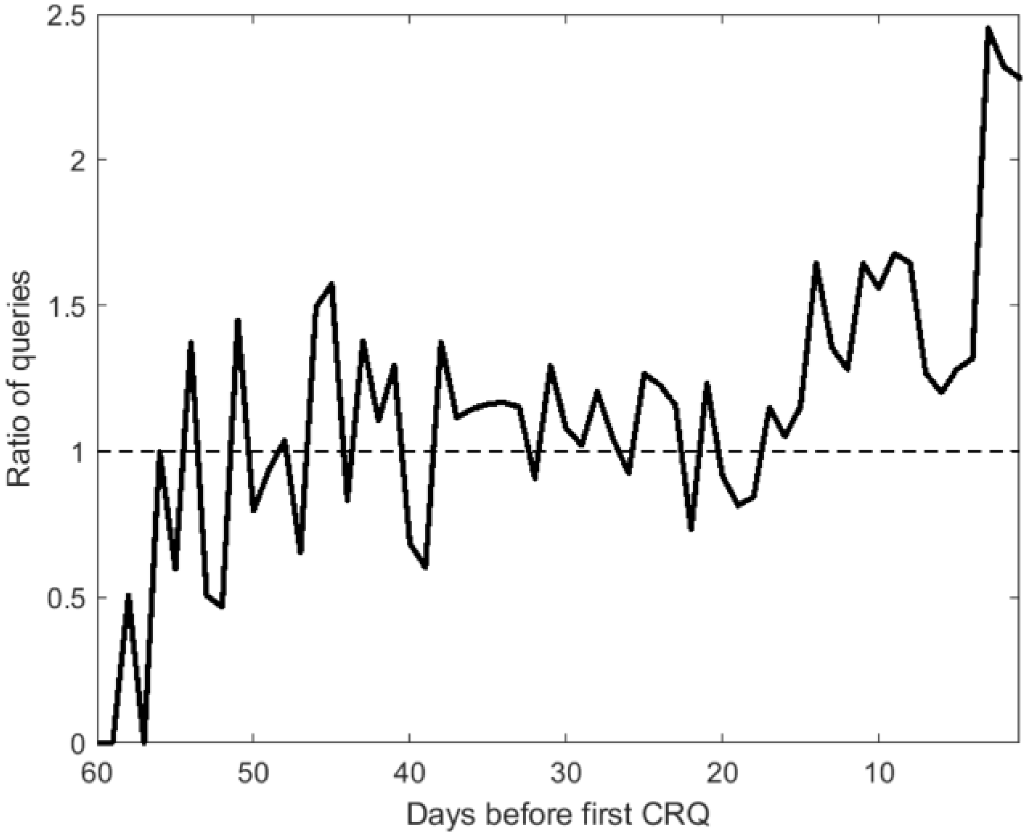


Table 2. Recipes and food ingredients that increased and those that decreased in the 14 days after the first celiac-related query, compared with the preceding 14 days.

Category	Increased	Decreased
Foods	Gluten-free pie crust	Honey cake
	Gluten-free pumpkin bread	Earthquake cake
	Gluten-free banana bread	Cucumber salad
	Gluten-free peanut butter cookies	Egg salad
	Gluten-free chocolate chip cookies	Pasta salad
	Gluten-free pancakes	Broccoli salad
	Roasted pumpkin seeds	Fish tacos
	Pumpkin soup	Ratatouille
	Cinnamon rolls	Tomato pie
	Pumpkin muffins	Tuna noodle casserole
Ingredients	Bean flour	All-purpose flour
	Brown rice flour	Dark rum
	White rice flour	Anisette
	Potato starch flour	Gin
	Rice flour	Serrano chile
	Xanthan gum	Peach schnapps
	Soy flour	Cherry
	Shortening	Pickles
	White sugar	Gelatin
	Ground walnuts	Sunflower seeds

Discussion

In this analysis of search engine queries, we found that symptoms known to be associated with celiac disease were searched for in the days preceding a first-time search for celiac disease or gluten. The symptom most likely to be queried more before than after a celiac disease query was diarrhea, a common clinical manifestation of celiac disease [21]. Though a variety of other symptoms and associated conditions were also queried before a CRQ, there was no combination of terms that resulted in an area under the curve of high discriminatory value. This lack of a discriminatory symptom set is in contrast to a prior analysis of this search engine investigating symptoms preceding a diagnosis of pancreatic cancer, which found that a set of search terms can identify 5% to 15% of patients with likely pancreatic adenocarcinoma while maintaining a low false-positive rate [13]. The lack of a consistent set of symptoms preceding interest in celiac disease or gluten-related disorders is congruent with a recent study evaluating medical records that found that clinical manifestations and associated diseases were largely ineffective at distinguishing patients with and without celiac disease [22]. Case finding (as opposed to population screening) is a widely accepted approach to identifying patients with celiac disease; however, a symptom-based approach appears to be unable to effectively distinguish patients with celiac disease from the general population, and this is borne out by our analysis.

The majority of patients with celiac disease now present without diarrhea and instead with other intestinal or extraintestinal symptoms [8]. Nevertheless, no single nonclassical manifestation is more common than diarrhea as a presenting feature, and thus, the plurality of patients have diarrhea, which might account for this symptom being the most likely to be mentioned before a CRQ [8]. Among the so-called nonclassical presentations, we found that queries for bloating and gastroesophageal reflux were associated with subsequent sustained queries for celiac disease. Among extraintestinal symptoms, headache, anxiety, depression, and attention-deficit hyperactivity disorder were associated with subsequent CRQs, raising the possibility that neuropsychiatric symptoms are a more prominent set of clinical features in celiac disease than is generally recognized. Patients with celiac disease have a greater risk of health care visits for headache both before and after celiac disease diagnosis [23]. Most studies have also found an association between both anxiety and depression and celiac disease [24], and the presence of depression appears to modify the relationship between adherence to the gluten-free diet and the severity of celiac disease–related symptoms [25]. Our findings suggest that these neuropsychiatric symptoms may be a prominent feature among individuals before they seek celiac disease testing.

In addition to the known intestinal and extraintestinal conditions and associated diseases, our analysis also yielded unexpected associations with subsequent CRQs, including cough, asthma, bleeding, influenza, itch, colitis, and Alzheimer disease (Table 1). Though cough and asthma are not thought to be a common manifestation of celiac disease, patients with celiac disease are somewhat more likely to have asthma [26], and several conditions that feature cough are also associated with celiac disease, including pneumococcal pneumonia and influenza [27,28]. Itch may be associated with celiac-specific queries because of dermatitis herpetiformis, a gluten-induced blistering rash that can be intensely pruritic [29]. Colitis may be associated with celiac-associated queries because of the known association between celiac disease and lymphocytic and collagenous colitis, 2 forms of microscopic colitis that may improve after the adoption of a gluten-free diet [30].

To our knowledge, this is the first study to analyze individual search engine data to identify antecedent symptoms in those who subsequently express an interest in celiac disease. A prior study analyzing regional patterns of Google searches for the gluten-free diet found that location-derived sociodemographic factors such as median income and proportion of residents who are non-Hispanic white were associated with an increased rate of searches for the gluten-free diet as compared with other diets [31]. In this study, we were able to analyze individual-level search data, allowing us to draw inferences about the variety of symptoms that precede awareness of the celiac disease or gluten as a possible underlying cause. The use of search engine queries allows us to evaluate symptoms that may be embarrassing for individuals to report to health care practitioners or on a traditional questionnaire [12]. Another strength of this study was its large sample size, encompassing over 6500 individuals who exhibited a sustained interest by performing a CRQ over more than 1 day.

This study also has a number of limitations. We are unable to distinguish diagnosed celiac disease from those individuals merely suspecting celiac disease. We attempted to mitigate against this by analyzing transient versus persistent interest, as those with persistent interest are more likely to have received a diagnosis of celiac disease; nevertheless, search engine analysis is unable to distinguish celiac disease from nonceliac gluten sensitivity. Though self-identified queries are rare, they can be useful to validate more widely used queries, and we did find that CRQs were highly correlated with a self-identified celiac query (eg, “I have celiac”) when compared with the general population; nevertheless, avoidance of gluten is far more common than diagnosed celiac disease [3]. Individuals with nonceliac gluten sensitivity have beliefs and attitudes that differ from those with celiac disease with regard to the health effects of gluten, the safety of genetically modified organisms, and other issues [32]. Our findings of multiple differences in queried recipes, including a rise in gluten-free baked goods, suggest that regardless of their celiac disease status, users with a CRQ are changing their dietary habits, at least in the short term. Regulations prohibit analysis of search engine query data beyond 18 months, which is considerably shorter than the latency period reported between symptom onset and celiac disease diagnosis (a mean of 11 years) in questionnaire studies [9].

In conclusion, in this analysis of celiac-related internet queries, we found an increase in antecedent searches for symptoms known to be associated with celiac disease such as diarrhea, bloating, and weight loss; a rise in searches for depression and anxiety; and an increase in symptoms that are associated with celiac disease but may not be reported to health care providers. We also found that the protean clinical manifestations of celiac disease are reflected in the diffuse nature of antecedent internet queries, underscoring the challenge of effective case-finding strategies. Future studies should investigate the unexpected associations found with CRQs in this study as well as the prevalence and natural history of neuropsychiatric symptoms in patients at the time of celiac disease diagnosis.

Conflicts of Interest

EYT is an employee of Microsoft, owner of Bing. BL has no conflicts of interest to disclose.

References

1. Lebwohl B, Sanders DS, Green PH. Coeliac disease. Lancet 2018 Dec 6;391(10115):70-81. [doi: 10.1016/S0140-6736(17)31796-8] [Medline: 28760445]
2. Rubio-Tapia A, Kyle RA, Kaplan EL, Johnson DR, Page W, Erdtmann F, et al. Increased prevalence and mortality in undiagnosed celiac disease. Gastroenterology 2009 Jul;137(1):88-93 [FREE Full text] [doi: 10.1053/j.gastro.2009.03.059] [Medline: 19362553]
3. Kim H, Patel K, Orosz E, Kothari N, Demyen MF, Pyrsopoulos N, et al. Time trends in the prevalence of celiac disease and gluten-free diet in the us population: results from the National Health and Nutrition Examination Surveys 2009-2014. JAMA Intern Med 2016 Dec 1;176(11):1716-1717. [doi: 10.1001/jamainternmed.2016.5254] [Medline: 27598396]
4. Biesiekierski J, Newnham E, Irving P, Barrett JS, Haines M, Doecke JD, et al. Gluten causes gastrointestinal symptoms in subjects without celiac disease: a double-blind randomized placebo-controlled trial. Am J Gastroenterol 2011 Mar;106(3):508-14; quiz 515. [doi: 10.1038/ajg.2010.487] [Medline: 21224837]
5. Davis W. Wheat Belly: Lose the Wheat, Lose the Weight, and Find Your Path Back to Health. New York: Rodale; 2011.
6. David P. Grain Brain: The Surprising Truth about Wheat, Carbs, and Sugar--Your Brain's Silent Killers. New York: Little, Brown And Co; 2019.
7. Blackett JW, Shamsunder M, Reilly NR, Green PH, Lebwohl B. Characteristics and comorbidities of inpatients without celiac disease on a gluten-free diet. Eur J Gastroenterol Hepatol 2018 Apr;30(4):477-483. [doi: 10.1097/MEG.0000000000001071] [Medline: 29315154]
8. Lo W, Sano K, Lebwohl B, Diamond B, Green PH. Changing presentation of adult celiac disease. Dig Dis Sci 2003 Feb;48(2):395-398. [doi: 10.1023/A:1021956200382] [Medline: 12643621]
9. Green P, Stavropoulos SN, Panagi SG, Goldstein SL, Mcmahon DJ, Absan H, et al. Characteristics of adult celiac disease in the USA: results of a national survey. Am J Gastroenterol 2001 Jan;96(1):126-131. [doi: 10.1111/j.1572-0241.2001.03462.x] [Medline: 11197241]
10. Paez MA, Gramelspacher AM, Sinacore J, Winterfield L, Venu M. Delay in diagnosis of celiac disease in patients without gastrointestinal complaints. Am J Med 2017 Dec;130(11):1318-1323. [doi: 10.1016/j.amjmed.2017.05.027] [Medline: 28623177]
11. Ludvigsson JF, Pathak J, Murphy S, Durski M, Kirsch PS, Chute CG, et al. Use of computerized algorithm to identify individuals in need of testing for celiac disease. J Am Med Inform Assoc 2013 Dec;20(e2):e306-e310 [FREE Full text] [doi: 10.1136/amiajnl-2013-001924] [Medline: 23956016]
12. Yom-Tov E. Crowdsourced Health: How What You Do on the Internet Will Improve Medicine. Cambridge, MA: MIT Press; 2016.
13. Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J Oncol Pract 2016 Dec;12(8):737-744. [doi: 10.1200/JOP.2015.010504] [Medline: 27271506]
14. Rosenblum S, Yom-Tov E. Seeking web-based information about attention deficit hyperactivity disorder: where, what, and when. J Med Internet Res 2017 Dec 21;19(4):e126 [FREE Full text] [doi: 10.2196/jmir.6579] [Medline: 28432038]
15. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using internet data. J Med Internet Res 2014 Jun 18;16(6):e154 [FREE Full text] [doi: 10.2196/jmir.3156] [Medline: 24943128]
16. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013 Jun 18;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
17. Yom-Tov E, Borsa D, Hayward AC, McKendry RA, Cox IJ. Automatic identification of web-based risk markers for health events. J Med Internet Res 2015 Jan 27;17(1):e29 [FREE Full text] [doi: 10.2196/jmir.4082] [Medline: 25626480]
18. Giat E, Yom-Tov E. Evidence from web-based dietary search patterns to the role of B12 deficiency in non-specific chronic pain: a large-scale observational study. J Med Internet Res 2018 Dec 5;20:e4. [doi: 10.2196/jmir.866] [Medline: 29305340]
19. Stork DG, Hart PE, Duda RO. Pattern Classification. New York: Wiley; 1973.
20. Ofran Y, Paltiel O, Pelleg D, Rowe J, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7:e45921. [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
21. Reilly NR, Fasano A, Green PH. Presentation of celiac disease. Gastrointest Endosc Clin N Am 2012 Oct;22(4):613-621. [doi: 10.1016/j.giec.2012.07.008] [Medline: 23083982]

22. Hujoel IA, Van Dyke CT, Brantner T, Larson J, King KS, Sharma A, et al. Natural history and clinical detection of undiagnosed coeliac disease in a North American community. Aliment Pharmacol Ther 2018 May;47(10):1358-1366. [doi: 10.1111/apt.14625] [Medline: 29577349]
23. Lebwohl B, Roy A, Alaedini A, Green PH, Ludvigsson JF. Risk of headache-related healthcare visits in patients with celiac disease: a population-based observational study. Headache 2016 May;56(5):849-858 [FREE Full text] [doi: 10.1111/head.12784] [Medline: 26972425]
24. Zingone F, Swift G, Card T, Sanders D, Ludvigsson J, Bai J. Psychological morbidity of celiac disease: a review of the literature. United European Gastroenterol J 2015 Apr;3:A. [doi: 10.1177/2050640614560786] [Medline: 25922673]
25. Joelson A, Geller M, Zylberberg H, Green P, Lebwohl B. The effect of depressive symptoms on the association between gluten-free diet adherence and symptoms in celiac disease: analysis of a patient powered research network. Nutrients 2018 Apr 26;10:E538. [doi: 10.3390/nu10050538] [Medline: 29701659]
26. Ludvigsson J, Hemminki K, Wahlstrom J, Almqvist J. Celiac disease confers a 1.6-fold increased risk of asthma: a nationwide population-based cohort study. J Allergy Clin Immunol 2011 Apr;127:1071-1073. [doi: 10.1016/j.jaci.2010.12.1076] [Medline: 21315431]
27. Simons M, Scott-Sheldon L, Risech-Neyman Y, Moss S, Ludvigsson J, Green P. Celiac disease and increased risk of pneumococcal infection: a systematic review and meta-analysis. Am J Med 2018 Jan;131:83-89. [doi: 10.1016/S0016-5085(17)32546-5] [Medline: 28801224]
28. Marild K, Fredlund H, Ludvigsson K. Increased risk of hospital admission for influenza in patients with celiac disease: a nationwide cohort study in Sweden. Am J Gastroenterol 2010 Nov;105:2465-2473. [doi: 10.1038/ajg.2010.352] [Medline: 20823839]
29. Reunala T, Salmi T, Hervonen K, Kaukinen K, Collin P. Dermatitis herpetiformis: a common extraintestinal manifestation of coeliac disease. Nutrients 2018 May 12;10:E602. [doi: 10.3390/nu10050602] [Medline: 29757210]
30. Stewart M, Andrews C, Urbanski S, Beck P, Storr M. The association of coeliac disease and microscopic colitis: a large population-based study. Aliment Pharmacol Ther 2011;33:1340-1349. [doi: 10.1111/j.1365-2036.2011.04666.x] [Medline: 21517923]
31. Laszkowska M, Shiwani H, Belluz J, Ludvigsson JF, Green PH, Sheehan D, et al. Socioeconomic vs health-related factors associated with Google searches for gluten-free diet. Clin Gastroenterol Hepatol 2018 Feb;16(2):295-297. [doi: 10.1016/j.cgh.2017.07.042] [Medline: 28826915]
32. Rabinowitz LG, Zylberberg HM, Levinovitz A, Stockwell MS, Green PH, Lebwohl B. Skepticism regarding vaccine and gluten-free food safety among patients with celiac disease and non-celiac gluten sensitivity. Dig Dis Sci 2018 May;63(5):1158-1164. [doi: 10.1007/s10620-017-4879-1] [Medline: 29243104]

Abbreviations

CRQ: celiac-related queries
QR: query ratio

Edited by G Eysenbach; submitted 11.12.18; peer-reviewed by F Montanaro, C Elena, R Verma, V Gianfredi; comments to author 31.01.19; revised version received 05.02.19; accepted 11.02.19; published 08.04.19.

Please cite as:
Lebwohl B, Yom-Tov E
Symptoms Prompting Interest in Celiac Disease and the Gluten-Free Diet: Analysis of Internet Search Term Data
J Med Internet Res 2019;21(4):e13082
URL: https://www.jmir.org/2019/4/e13082/
doi:10.2196/13082
PMID:30958273

Original Paper

Using a Facebook Forum to Cope With Narcolepsy After Pandemrix Vaccination: Infodemiology Study

Karin Blomberg[1], RN, PhD; Mats Eriksson[1], RN, PhD; Rickard Böö[2], BSc; Åke Grönlund[2], PhD

[1]School of Health Sciences, Örebro University, Örebro, Sweden

[2]School of Business, Örebro University, Örebro, Sweden

Corresponding Author:
Karin Blomberg, RN, PhD
School of Health Sciences
Örebro University
Faculty of Medicine and Health
Örebro, S-70182
Sweden
Phone: 46 19 301270
Email: karin.blomberg@oru.se

Abstract

Background: In 2010, newly diagnosed narcolepsy cases among children and adolescents were seen in several European countries as a consequence of comprehensive national vaccination campaigns with Pandemrix against H1N1 influenza. Since then, a large number of people have had to live with narcolepsy and its consequences in daily life, such as effects on school life, social relationships, and activities. Initially, the adverse effects were not well understood and there was uncertainty about whether there would be any financial compensation. The situation remained unresolved until 2016, and during these years affected people sought various ways to join forces to handle the many issues involved, including setting up a social media forum.

Objective: Our aim was to examine how information was shared, and how opinions and beliefs about narcolepsy as a consequence of Pandemrix vaccination were formed through discussions on social media.

Methods: We used quantitative and qualitative methods to investigate a series of messages posted in a social media forum for people affected by narcolepsy after vaccination.

Results: Group activity was high throughout the years 2010 to 2016, with peaks corresponding to major narcolepsy-related events, such as the appearance of the first cases in 2010, the first payment of compensation in 2011, and passage of a law on compensation in July 2016. Unusually, most (462/774, 59.7%) of the group took part in discussions and only 312 of 774 (40.3%) were lurkers (compared with the usual 90% rule of thumb for participation in an online community). The conversation in the group was largely factual and had a civil tone, even though there was a long struggle for the link between the vaccine and narcolepsy to be acknowledged and regarding the compensation issue. Radical, nonscientific views, such as those expounded by the antivaccination movement, did not shape the discussions in the group but were being actively expressed elsewhere on the internet. At the outset of the pandemic, there were 18 active Swedish discussion groups on the topic, but most dissolved quickly and only one Facebook group remained active throughout the period.

Conclusions: The group studied is a good example of social media use for self-help through a difficult situation among people affected by illness and disease. This shows that social media do not by themselves induce trench warfare but, given a good group composition, can provide a necessary forum for managing an emergency situation where health care and government have failed or are mistrusted, and patients have to organize themselves so as to cope.

(J Med Internet Res 2019;21(4):e11419) doi:10.2196/11419

KEYWORDS

narcolepsy; mass vaccination; social media

Introduction

Background

In 2010, newly diagnosed narcolepsy cases among children and adolescents were reported in several European countries as a consequence of comprehensive national vaccination campaigns with Pandemrix against H1N1 influenza (swine flu) that took place during the winter of 2009–2010 [1]. In Sweden, approximately 350 young persons acquired narcolepsy after vaccination against swine flu [2,3]. Narcolepsy is a neurological disease that involves a disturbed regulation of wakefulness and sleep. The main symptoms are severe insomnia and daytime sleep attacks that cannot be controlled. Cataplexy, a sudden loss of muscle tone and a feeling of paralysis, is commonly associated with narcolepsy and can have a significant impact on the daily lives of the affected persons. There is no curative treatment, only symptom relief with stimulant drugs, antidepressants, and sodium oxybate (Xyrem), which all have fairly severe adverse effects [4].

Despite a lack of scientific documentation of the life situations of young persons who acquired narcolepsy after swine flu vaccination, several personal descriptions have been published in the media (newspapers and the internet) and aired on television. Having grown up in the digital age, young people today are accustomed to expressing their opinions quickly and widely on social media. In a crisis, social media could potentially be a significant means for disseminating and collecting information and could improve emergency management, for example. On the other hand, social media have been ascribed the power to misinform [5]. Social media are described as an amplifier of opinions more than as an arena for objective discussions [6-8], and the health information presented is often inaccurate or not in line with official recommendations for prevention and treatment. For many people, attitudes toward vaccination are shaped not just by health care but also by other information sources published on websites and social media. In the case of vaccination campaigns, several antivaccination movements opposed to, for example, vaccination against human papillomavirus and measles have been started through social media [9,10]. The relation between narcolepsy and the swine flu vaccination campaign is often used in blogs as an example of why the medical authorities should not be trusted [11].

As social media and the internet are seen by many as a primary source of health-related information [12,13] and can change how young people share information and make decisions regarding their health and well-being, it is important to investigate how information, perceptions, and attitudes are spread. The power of social media puts new demands on communication strategies of health care systems.

The swine flu virus was first isolated from pigs in 1931. The first known death of a human due to the virus was described in 1976. In 2009, a new type of swine flu virus emerged in California, USA, which was quite harmless to most people, whereas some individuals who did not belong to traditional risk groups became extremely ill. Within a few months, the flu became pandemic, and many countries decided to offer their citizens a pandemic vaccine campaign [1]. In connection to the 2009–2010 pandemic, about 60% of Sweden’s population were vaccinated against swine flu [14,15]. In 2010 came the first reports from Finland and Sweden of narcolepsy among children and adolescents vaccinated with one of the vaccination types, Pandemrix. Since then, studies from Sweden, Finland, the United Kingdom, and Ireland have demonstrated a link between narcolepsy and the Pandemrix vaccine, with this vaccine producing a 3-fold increase in the risk of narcolepsy [1]. When the link between the vaccine and narcolepsy was clarified, claims were made by the families of the affected children, and later also by the adolescents themselves, for compensation from pharmaceutical companies, medical insurers, health authorities, and others. In Sweden, health care, including disease prevention, is regulated and performed by the public sector; consequently, the government was the main target of these demands. Among other things, discussions revolved around the maximum limit of compensation. Table 1 lists the most significant events during the period from March 2009, when the virus was detected, to July 2016, when the dispute in Sweden was settled by legislation regulating compensation to affected individuals. Different actions to call for higher compensation were initiated, using channels such as newspapers, television, and the internet.

Table 1. Significant events concerning the swine flu outbreak, vaccination, narcolepsy, and compensation regulation.

Date	Event
March 2009	Reports emerge of a new type of swine flu in California and Texas and spreading to Mexico.
June 2009	The World Health Organization declares a swine flu pandemic. Vaccine development begins.
October 2009	The Swedish vaccination campaign starts.
August 2010	Reports of narcolepsy in vaccinated children emerge from Finland and Sweden.
July 2011	The European Medicines Agency recommends limited use of Pandemrix in people <20 years of age.
October 2011	The first persons who had narcolepsy after vaccination receive compensation from the Swedish pharmaceutical insurance industry.
March 2013	The first register study is published, showing the link between Pandemrix and narcolepsy, as also exposed in the daily media.
May 2015	The Swedish government sets the maximum compensation amount for lost income at SEK 10 million.
July 2016	A new law is passed regarding compensation from the Swedish pharmaceutical insurance industry to affected persons who experienced the first symptoms of narcolepsy <24 months after vaccination.

Objective

The aim of this study was to examine how information was spread, and how opinions and beliefs about narcolepsy as a consequence of Pandemrix vaccination were formed through discussions on social media. For this purpose, we examined a series of messages posted on social media, as well as connections among people participating in the discussions. One research question was whether, in general, social media groups for persons affected by narcolepsy were factual and constructive and helped those with narcolepsy to cope with the situation, and whether the negative campaigning against vaccination came from other sources. If this were the case, then social media were not the main driver of nonfactual discussions; rather, group composition was.

This study was part of a project investigating the life situations of narcolepsy-affected young persons, and the use and meaning of social media and the internet as a means of support and communication of opinions [2]. We hoped that the results of the study would increase knowledge of how the use of social media can affect trust in health care and attitudes toward future health campaigns, and thereby assist in developing interventions to support increased trust in and compliance with such campaigns.

Methods

Identification of Discussion Forums

This study was a descriptive retrospective analysis of a series of messages posted on social media, as well as connections between users, between September 2010 and July 2016. We conducted the study in 2017. We found 18 Swedish discussion forums on the Web that concerned narcolepsy and that had started around 2010. As Table 2 shows, in most of these, the volume of activity was very low, with only a few posts. Most of the forums also had no posts from recent years: the Pandemrix vaccination took place in 2010, the first cases of narcolepsy occurred the same year, and most posts in these groups were from around that time. As we were interested in online discussions, where people can meet, take part on equal terms, and go in-depth into issues they find important, some of the sites in Table 2 were irrelevant. Numbers 15 to 18 were blogs, which means they were not really discussion forums. Even though some blogs allow comments, only the blog owner can make posts. Number 17, YouTube, also is not a discussion forum but a publication site. We also considered Twitter (number 11) to be irrelevant because Twitter is by its format not suitable for lengthy discussions, and possibly for that reason, there were no discussions, only a few disconnected tweets.

We selected the only group that had a large number of threads and where activity had been high throughout the 6 years—on average 23 posts per month (number 1). This is a Facebook group in Sweden named “Narcolepsy after the Pandemrix Vaccine,” which had 774 members as of September 12, 2016.

The second most active group was Flashback with 250 threads, or about 3 per month on average. We considered that a quite low activity, in particular because many discussions were offensive, promoted conspiracy theories, or were not about narcolepsy but about entirely different topics, such as fluoride-free toothpaste. Textbox 1 shows some examples (translated from the Swedish language).

The Facebook group is presented as “a group for us, or relatives, who have the diagnosis of narcolepsy or have similar symptoms like narcolepsy/sleeping sickness after vaccination.” Although the group is open, it is reasonable to assume that most people in the group are either affected or close relatives of those affected by the disorder. When the group started, a total of 200 people were affected by, that is, had a diagnosis of, narcolepsy after vaccination. Adding a few who had the symptoms but did not yet have a diagnosis, and between 2 and 3 relatives per affected person, brings the total to around 774. This suggests that the Facebook group was (voluntarily) limited to narcolepsy patients and their close relatives. The group has been active since September 2010, and at the time of our investigation there were 1671 posts, which had generated a total of 10,906 comments.

Table 2. Forums found by searching for “narcolepsy,” September 12, 2016.

Forum		Web address	Number of threads up to September 12, 2016	Activity
1	Narcolepsy after the Pandemrix vaccine	https://www.facebook.com/groups/122068704510686/	1671	High and regular
2	Flashback	https://www.flashback.org/sok/Narkolepsi	250	Sporadic and not always about narcolepsy
3	Fragbite	http://fragbite.se/search/?q=Narkolepsi&t=forum	3	Sporadic
4	Allt för föräldrar	http://www.alltforforaldrar.se/snack2/search.php?searchid=357008	59	Low
5	NeuroFörbundet	http://neuroforbundet.se/l.aspx?u=http://neuroforbundet.se/community/grupper/forum/?clubId%3d8	Requires login	N/A[a]
6	Hamsterpaj	http://www.hamsterpaj.net/soek/#!/000100/narkolepsi	1	Low
7	Vetenskap och folkbildning	http://forum.vof.se/index.php	N/A	N/A
8	MEF	http://me-cfs.se/mef-forum/index.php?topic=662.msg3466#msg3466	1	Low
9	Netdoktor.se	http://www.netdoktor.se/neurologi/diskussioner/	1	Low
10	Fuska.se	http://fuska.se/forum/index.php?app=core&module=search&do=search&fromMainBar=1	4	Sporadic
11	Twitter (#Narkolepsi)	https://twitter.com/hashtag/Narkolepsi https://twitter.com/search?q=Narkolepsi&src=typd	No discussions were found, only sporadic unconnected tweets	N/A
12	Narkolepsiföreningen (Facebook)	https://www.facebook.com/narkolepsiforeningen/	N/A	Low
13	Doktorn.com (Facebook)	https://www.facebook.com/doktornpunktcom/	N/A	Low
14	YouTube	https://www.youtube.com/watch?v=d7HGfPrtlkY	N/A	Low
15	Tankebrott blogg	https://tankebrott.nu/2010/08/24/orsak-verkan-narkolepsi-och-pandemrix-och-frustration/	N/A	High
16	Matildas blogg	http://dagsattvakna.for.me/narkolepsi.html	N/A	Low
17	Rune Lanestrands blogg	https://runelanestrand.wordpress.com/tag/narkolepsi/	N/A	Low
18	Svensson blogg	http://blog.zaramis.se/tag/narkolepsi/	N/A	Low

[a] N/A: not available.

Textbox 1. Examples of threads on Flashback.

- “A friend told [me] that vaccines that we get as kids make us weird, like gay and other stuff.”
- “I read that some kids get autism and other shit.”
- “If you can get narcolepsy from vaccine you can in principle imagine that you can for example become gay.”
- “Vaccines generally give more diseases. More and more people begin to understand that vaccination has become one of the most profitable activities for the pharmaceutic companies. There are many indications that they manipulate statistics, and when you look at it the way you should, vaccines have no or very poor effect. All statistics I have seen suggest that.”
- “‘BLOODY CREEPS.’ I am fucking boiling! How the hell can they have the stomach to complain about that [government compensation to people who got narcolepsy from a vaccine]. They can retire and not do any work for the rest of their lives. I fucking think that those who complain should lose all their compensation!”

Data Collection

We collected some data directly from the Facebook group, such as a count of posts (a “post” is the start of a discussion on Facebook) and comments. For the content analysis, we selected posts with more than 10 comments from months with high activity, defined as 200 or more posts or comments per month. This selection criterion yielded 141 posts out of the total of

1671 that were available at the time. The reason for this selection was that we wanted to see what the discussions that had "caught on" were about. Posts that generated many comments could be considered to have been more interesting to the group than those that do not. While the limit of 10 comments was somewhat arbitrary, it weeded out posts that received only marginal comments (eg, "good to hear from you"), corrections (eg, "sorry, 25 is the correct number"), or clarifications (eg, "Where did you hear that?" "In...;" "But they made an update this morning saying..."). The selection includes 141 of the 1671 (8.4%) total number of posts and 3086 of 10,906 (28.3%) comments. The remaining 1530 posts (that we did not select) had an average of 5 comments.

We downloaded text data from the Facebook group using Netvizz v1.31 [16] to identify posts with more than 10 comments. We used NVivo 11 for Mac (QSR International) for quantification of data (ie, analyzing the numbers of posts and users).

Data Analysis

Text data were sorted by the third author (RB) into posts and comments regarding users. We used an inductive approach, which involved reading the data and identifying headings describing the content of the posts and comments. The analysis was conducted in several iterations involving all authors, and including discussion of interpretations of what was said in posts and comments, and various alternatives for quantifying the results, for example by author, keywords, and frequency. The analysis process was driven by a set of analytical questions regarding content, participation, and communication style. The content questions were (1) What are the most common discussion topics? (2) How do discussion topics change over time? and (3) What triggers the major discussions (eg, external events such as news media stories or government or health care system actions, or internal events such as somebody working to raise awareness or personal experiences shared)? The participation questions were (1) How many people or how large a share of group members participate in discussions? and (2) How many people *start* discussions? The question pertaining to communication style was What is the general tone of the communication (positive—negative, factual—polemic, personal—general)?

Ethical Considerations

As mentioned above, the Facebook group is open and all posts and comments, as well as the names or pseudonyms of the people who have written them, are publicly visible. In this study, no individual's post or comment in the group could be identified, either by group member name or by the content of their posts. We handled data in line with the Declaration of Helsinki [17]. According to Swedish law, studies on social media do not require ethical approval; however, the overall project was approved by the Regional Ethical Review Board of Uppsala (registration no. 2013/505).

Results

Content

The Facebook group activity was fairly high throughout the years 2010 to 2016, although there was considerable monthly variation (Figure 1). The peaks of activity corresponded to the major events over the years, that is, the first cases of narcolepsy in late 2010, the first compensation paid in late 2011, and the passage of a law on compensation in July 2016 (see Table 1). But there was also sustained high activity from late 2012 to early 2014. While no milestone event took place during that period, it was a time when new cases of narcolepsy were being discovered and many more people found themselves struggling with a difficult situation.

Figure 1. Posts and comments over time.

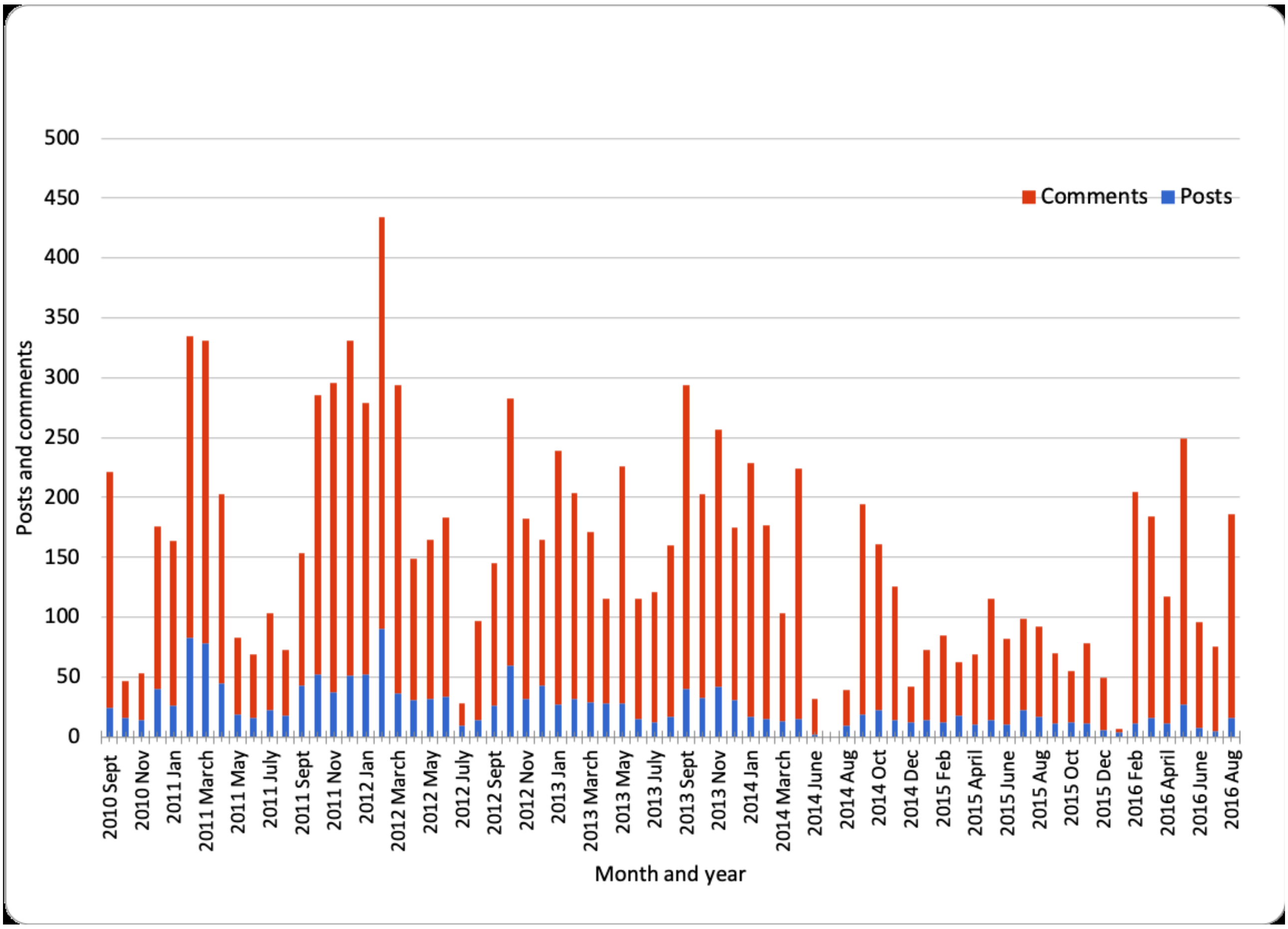


Table 3. The most common discussion topics.

Rank	Topic	Number of posts
1	Adverse effects of Pandemrix or the vaccination	37
2	Drugs and medication	28
3	Narcolepsy diagnosis	17
4	Insurance and compensation	16
5	Doctor or hospital visits and treatments	12
6	Vaccination in general	8
7	Events and meetings	8
8	Upset discussions about public statements by an official or journalist	5[a]
9	Scientific research	2
10	Other	8[b]

[a] This includes statements by radio hosts (n=2), politicians (n=2), and the Head of the Swedish Institute for Infectious Disease Control (n=1).

[b] These include 8 different topics with only 1 post each, ranging from Christmas greetings to blood donation to a conspiracy theory about whether the government had known about the adverse effects.

Changes in Discussion Topics Over Time

Table 3 lists the most common discussion topics.

Discussion topics changed over time (Figures 2-7). Some topics, such drugs and medication (Figure 3) or narcolepsy diagnosis (Figure 4), were on the agenda more or less all the time, albeit with varying frequency. Peaks corresponded to significant narcolepsy events, such as, initially, the increase in cases in 2011–2012, then the corroborated link between the vaccination and the disease in 2013. Adverse effects also remained a hot topic for a long time, peaking in late 2012 and 2013 when the relation to the vaccination was first investigated and later established (Figure 4). Other themes, such as insurance and compensation (Figure 6) and doctor or hospital visits and treatments (Figure 7), emerged frequently for shorter periods, following significant events after the vaccination campaign had started in 2010 (see Table 1). Doctor or hospital visits was a frequent topic during 2011–2012 when the pandemic exploded. It was followed by another, smaller peak in May 2013, which coincided with, and could be related to, increased publicity at the time about the link between Pandemrix and narcolepsy, based on the first Swedish registry study [14,15] published on the website of the Swedish Medical Products Agency in 2013 and exposed in the daily media. The topic was then related to the importance of having a diagnosis in order to qualify for compensation. The topic narcolepsy diagnosis was a frequent discussion topic from 2013, when the cause was established and discussions on compensation became active in the media, until 2016, when the new law settled the compensation issue. Insurance was another major topic in late 2013, when the first study established the link between Pandemrix and narcolepsy, which naturally triggered discussion about compensation.

Even the more constant topics had ups and downs over the investigated period. For example, discussions on the adverse effects of Pandemrix were frequent throughout the period, but there was a major peak from February 2012 to January 2013, coinciding with discussion in the media during that period (Figure 4).

Figure 2. Post topics with high activity.

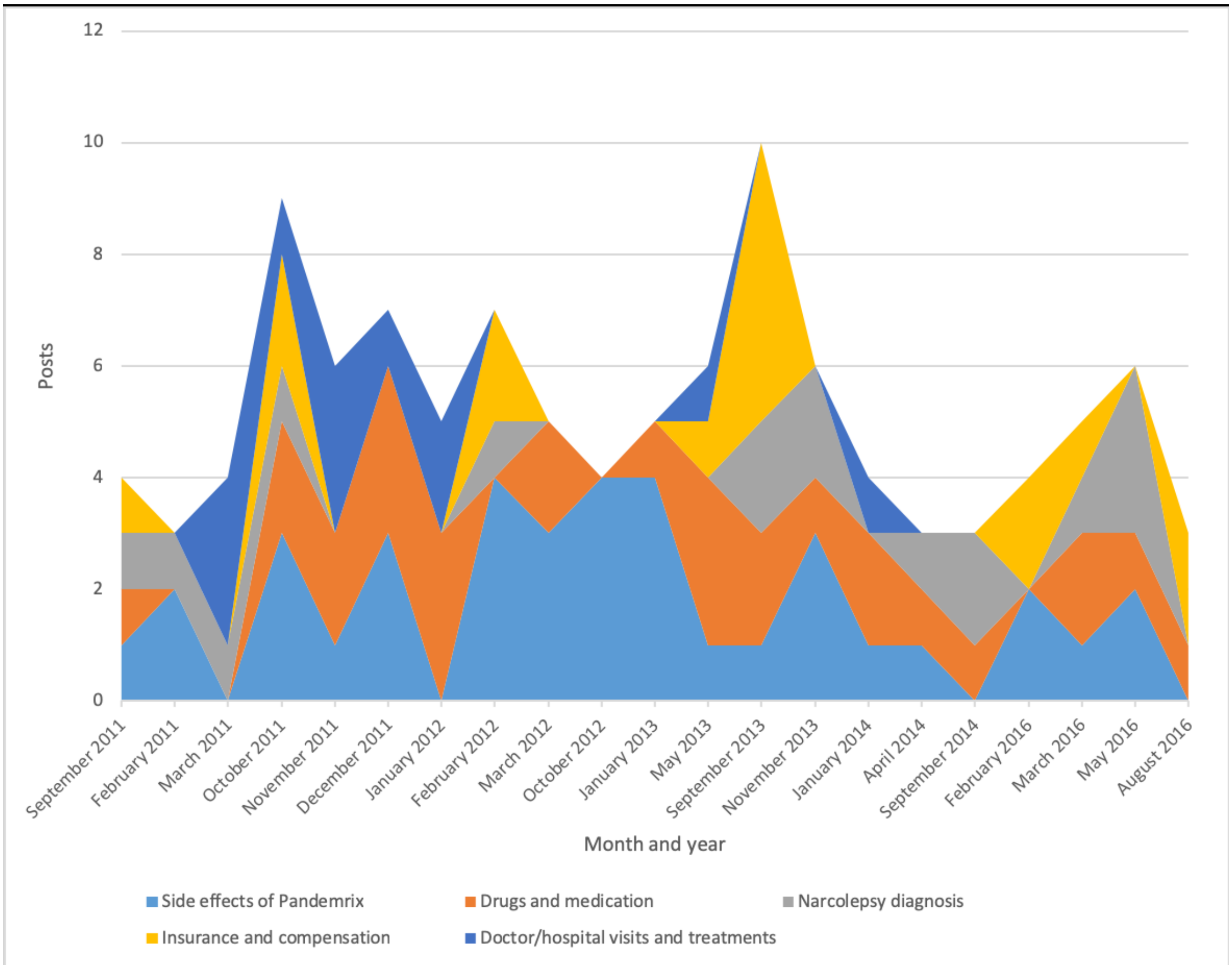


Figure 3. Posts on drugs and medication.

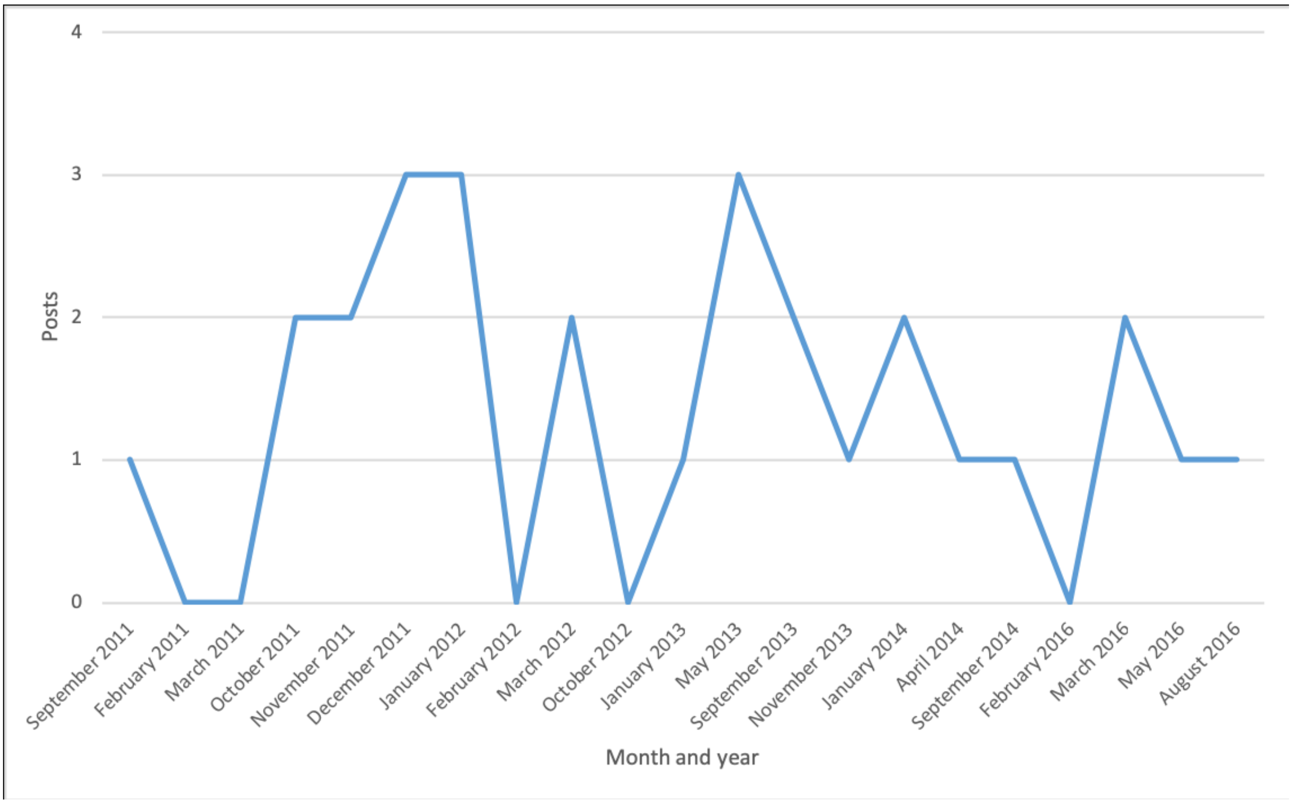


Figure 4. Posts about the adverse effects of Pandemrix.

Figure 5. Posts about narcolepsy diagnosis.

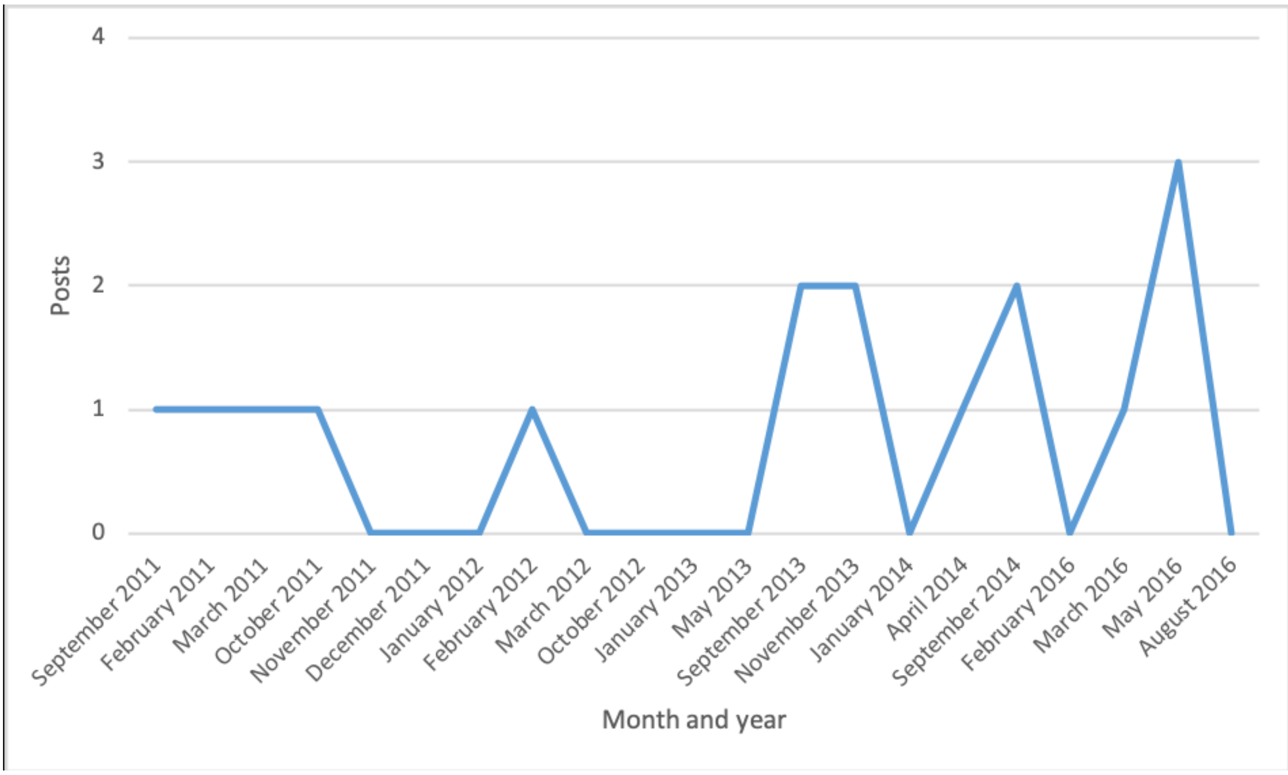


Figure 6. Posts about insurance and compensation.

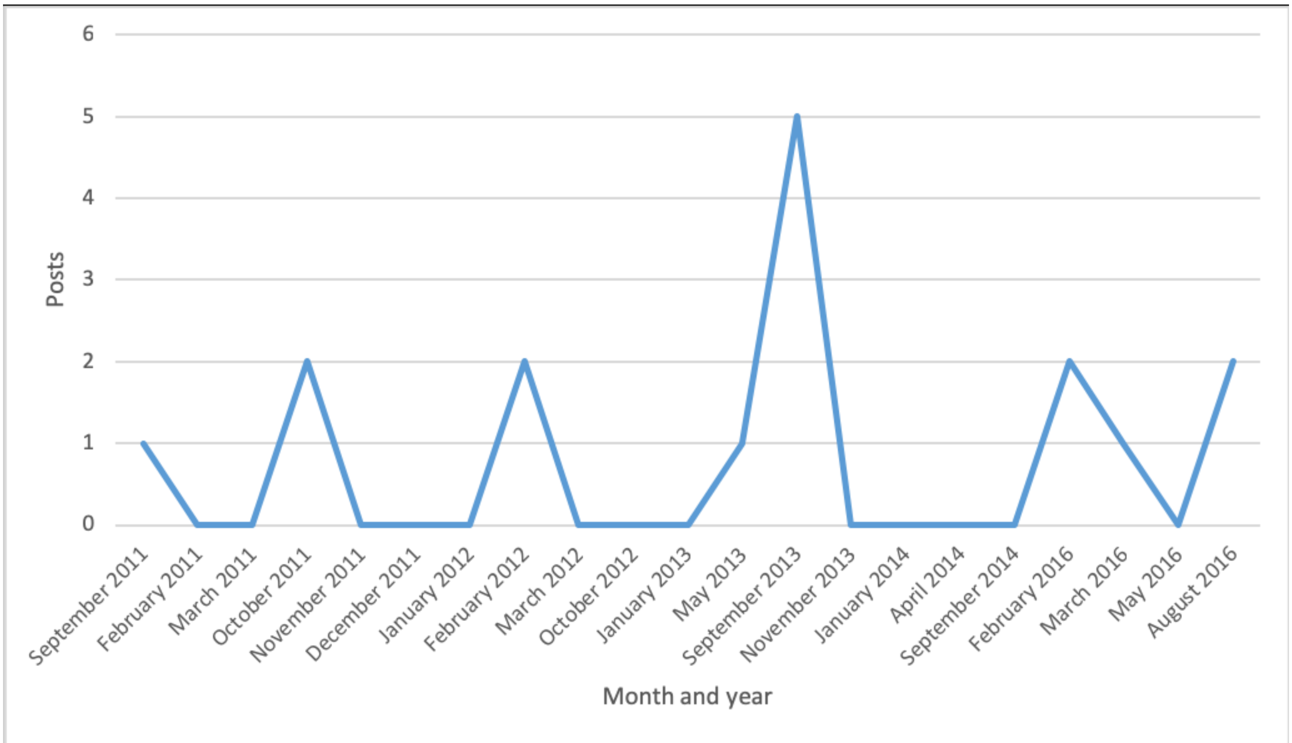


Figure 7. Posts about doctor or hospital visits and treatments.

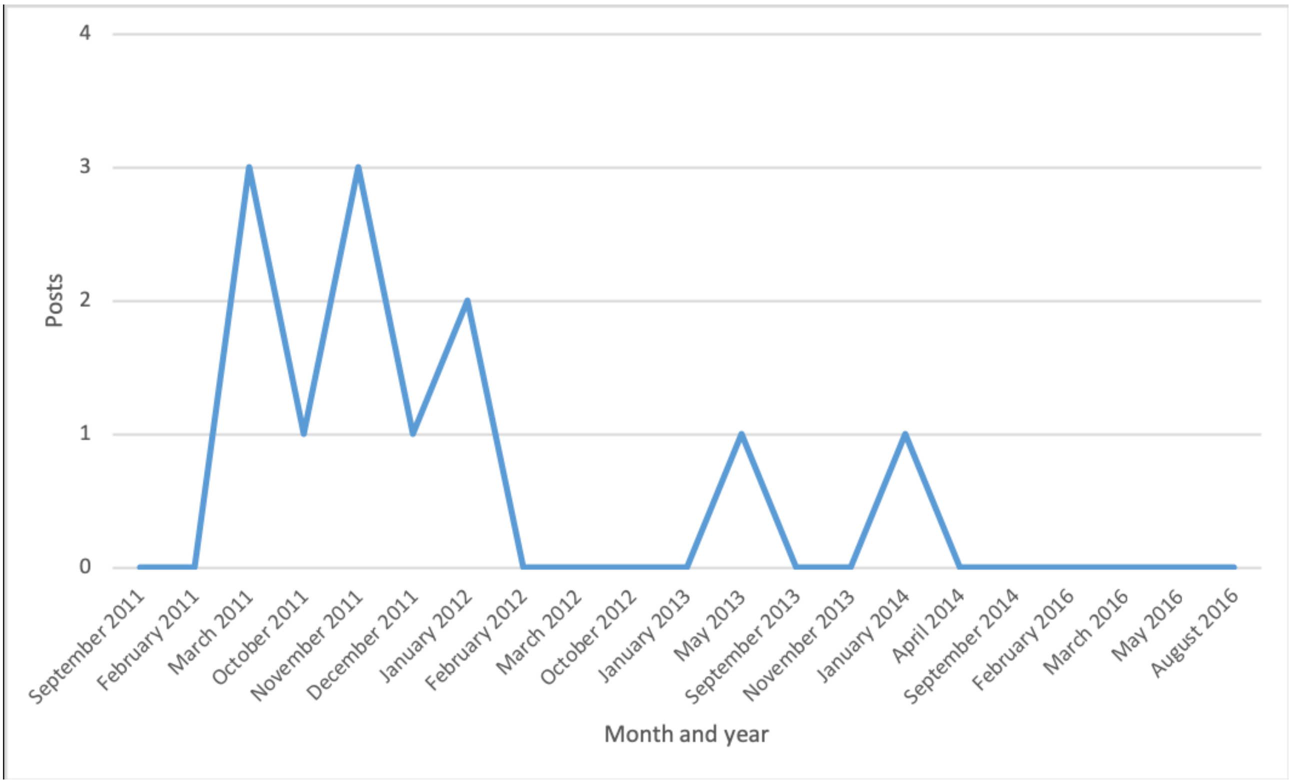


Table 4. Rationale behind the posts.

Rationale	Number of posts
Searching for information or others' experience	46
Sharing personal experience or information	44
Sharing information from external sites	27
Sharing information about an upcoming event or meeting	8
Airing statements by the media	8
Other	7
Presenting a conspiracy theory	1
Total	141

Discussion Triggers

While peaks in discussion topics can be related to significant events in the history of the vaccination and the ensuing pandemic, there is still the question of which events, situations, needs, or emotions triggered specific major discussions. As Table 4 shows, the most common cause by far prompting the writing of posts was a personal need for finding and sharing information. Only a small number of posts (n=8) contained "negative" discussions. These were reactions to public statements by officials or journalists that were felt to be derogatory. Only 1 post aired a conspiracy theory, and this did not take root.

The general tone of the discussions was factual, friendly, and positive despite the problematic, to say the least, situation many of the group members found themselves in. They tried to share information and experiences with each other in the hope of easing the life of "the victims." The posts were generally by a member seeking answers and experiences from the other members of the group, or someone wanting to share information or personal experiences with the group.

Of the 141 posts, only 13 (9.2%) had a more negative character, and even in these threads the tone was mainly factual. When the members expressed anger in their posts, it was anger they felt either toward individuals (politicians, officials, or a media host) who had made statements about the situation that group members felt were unfair or derogatory (n=4) or toward individuals who had purportedly exploited the situation to commit insurance fraud (n=1), or the posts expressed dissatisfaction with a doctor (n=2) or anger toward the Head of the Swedish Public Health Agency (n=1). Alternatively, it was about insurance and compensation (n=2), about incorrect information on a website (n=1), or demands for a public apology from the government (n=1). In one case a conspiracy theory claiming that the government had known about the adverse effects beforehand was posted.

Participation

The Facebook group had 774 members by the time of our investigation. Over the group's lifetime to that point, a total of 1671 posts had been published, with 10,906 comments. We selected the major discussions: those that generated 10 or more comments.

In all, 59.7% (462/774) of the group members had participated in the discussions by either creating or commenting on a post. Slightly more than one-third of the group members (277/774, 35.8%) had created a post (that is, started a discussion).

The 5 people who contributed the largest number of discussions accounted for 26.0% (435/1671) of the posts published in the group. They also provided 22.2% (2423/10,906) of all the comments on the forum.

The 10 people (10/774, 1.3% of the population) who created the largest number of discussions accounted for 42.19% (705/1671) of the posts that generated the most comments. They also provided 35.19% (3838/10,906) of all comments on the forum. The single most active member had created 107 posts, or 6.4% of the total 1671 posts. A total of 7 members had created 50 or more posts each, and 32 members had created more than 10 posts. All the other 445 active posters had provided an average of 1.7 posts.

These participation rates are considerably higher than the general rule of thumb for social media use. The "90%-9%-1% rule" [18] states that, in most online communities, 1% of users account for almost all the action, 9% of users contribute a little, and 90% of users are lurkers who never contribute. In the community we analyzed, we found agreement on the last statement of the rule but not the first or second. In this population, 1.2% (9/774) of participants accounted for 41.8% (281/671) of the posts and 35.5% (3871/10,906) of the comments. While not statistically constituting the majority, this is certainly a major part. We also found that the majority of members of the community—59.0% (457/774) if we take away the most active 1%—contributed at least something. The share of lurkers, if defined as people who never actively participate, in our study was then only 40.1% (310/774). Even though many had made only 1 or a very few posts, posting means starting, or trying to start, a discussion, which can be considered more active participation than commenting on someone else's post.

Generally speaking, the Facebook group we studied can be considered fairly participatory by general social media standards, even though a few people clearly dominated the discussions. Presumably this means not only that participants considered the group to be highly relevant, but also that active participation was important, and that many people trusted the group enough to share even sensitive stories from their lives.

Discussion

Principal Findings

The aim of this study was to examine how information was disseminated, and how opinions and beliefs about narcolepsy as a consequence of Pandemrix vaccination were formed by discussions on social media. To investigate this, we examined a series of messages posted on social media over a 6-year period. The focus was on people affected by narcolepsy after having been vaccinated against swine flu. Out of 18 Swedish groups originally discussing this topic, the Facebook group we examined was the only one that has remained active throughout the years.

Our research question, whether social media groups for affected patients are generally factual and constructive and help the patients cope with the situation and, further, whether negative campaigning against vaccination comes from other sources, was motivated by a more general research interest. We assumed that social media do not inherently drive nonfactual discussions or less honest discussion styles but, rather, that social media are a useful tool not just for social purposes but also for really difficult discussion on very serious matters that concern people's health. If this assumption holds, then social media can be very useful not just for the people participating but also for health care providers. Interventions using social media, such as Facebook, in health care are increasing—for example, interventions for awareness of breast cancer [19], physical activity promotion among adolescent and young adult childhood cancer survivors [20,21], sexual health promotion [22], HIV communication [23], and enhancement of positive health outcomes among adolescent and young adults [24]: all studies referenced here indicated positive benefits. However, there is still a lack of robust evidence of effects and concerning how to best design such interventions. Despite an increased use of social media by health care providers, issues remain concerning how to best provide health information and support that are trustful and that promote healthy behavior among people. The Facebook group studied here was self-organized, and this may have been a factor affecting people's trust in it.

If social media themselves do not drive nonfactualism and asocial behavior, then certainly people and group composition might. This particular group was composed of people who shared the same serious situation and great need, not just for information, but also for advice and the possibility to share their concerns with others. It should be noted, however, that most of them did not know each other in real (offline) life. They lived in different cities and were dispersed across Sweden. In this respect theirs was a truly virtual community.

From previous research on social media we know that distance makes social ties weaker. Harsh language is more common when people live far apart from each other than when they live in the same city and there is at least a theoretical chance that they might physically meet. It appears that in the case we studied, the shared situation served the purpose of making people feel close enough to make the community "real." This was supported, for example, in studies of a group for persons with Huntington disease, where exchanging informational and emotional support was a key function [25], and a group for persons with amyotrophic lateral sclerosis that served as a source for distributed knowledge [26].

This type of scenario creates opportunities for health care, as many patients with some serious condition share the same situation: they need to learn about their problem and possible treatments, and they have to cope with their situation. They often do not know people with the same condition at the outset but want to find them as the need arises. Whatever the condition, in most cases most people with the same condition do not live next door. Some people prefer to share and disclose experiences of illness and health in forums such as Facebook, while others are hindered by the lack of anonymity in social media [27]. The resource we presented might also be limited to people with internet skills and access, and may exclude people from socially disadvantaged groups with lower socioeconomic status [28].

Of course, 1 case does not prove a hypothesis, but at least this case gives positive evidence of the possibilities that social media can present to these people living with an uncommon but serious condition in a difficult and uncertain situation. Despite several years having passed since the onset of the first vaccination-related cases of narcolepsy, this Facebook group still exists. This could indicate a need for support—support from society and, even more important, from health care services.

Although the discussions in the group were generally factual, positive, and directed toward problem solving and coping, rather than bitterness and conflict, we did find one post putting forward a conspiracy theory. While there may have been more negativity in the posts with fewer than 10 comments, these did not lead to any longer discussion. Much has been written about how social media amplifies opinions [6-8], but it has also been found that conflict in discussion forums often occurs between groups rather than within them. For example, at “hate sites,” traditional media are criticized while criticism against hate sites mainly occurs at traditional media sites. The same seems to be the case here. There was little conflict in this group, either internally or externally (eg, through members attacking people outside the group), but it is easy to find other places on the Web where antivaccination views are vigorously and often aggressively expressed. A systematic review of 42 studies, 16 of which explored Facebook posts, concluded that most had a beneficial or neutral impact on the clinical outcome of chronic diseases [29].

It would be interesting to investigate why some posts did not generate comments—was it because they were inappropriate, marginal, or made by people who were not centrally positioned in the group? This investigation, however, was not about the social situation of the group per se, but rather about the general role of a group such as this, regarding discussions about a severe disorder.

Conclusion and Implications

The aim of this study was to examine how information was shared, and opinions and beliefs about narcolepsy as a consequence of Pandemrix vaccination were formed by discussions on social media. To investigate this, we examined a series of messages posted on social media over a 6-year period. We found, first, high group activity throughout the years 2010 to 2016, with peaks corresponding in time to major narcolepsy-related events, such as the appearance of the first cases in 2010, the first compensation paid in late 2011, and passage of a law on compensation in July 2016. Second, unusually, a majority (about 60%) of the group members took part in discussions and only 40% were lurkers (in contradiction to the 99%-9%-1% rule of thumb for participation in an online community). Third, the conversation in the group was largely factual and had a civil tone, even though there was a long period of struggle to get acknowledgement of the link between the vaccine and the disease and regarding the compensation issue. Fourth, radical, nonscientific views, such as those of the antivaccination movement, did not shape the discussions in the group but were active elsewhere on the internet.

The Facebook group we studied is a good example of social media use for patient self-help in a difficult situation. The example shows that social media do not by themselves induce trench warfare but, given a good group composition, can provide a necessary forum for managing an emergency where health care and government have failed and are mistrusted, and patients have to organize themselves so as to cope.

The critical factor is not social media use, but group composition. The Facebook group studied here appeared to have consisted only of people directly concerned—individuals directly affected and their close families, and, furthermore, people who appeared to believe in facts and science, even though the pandemic was started by a health care mistake. This socially coherent and fact-oriented discussion group survived 6 years, while 17 other groups, many of which included much more confrontational language and views, that appeared at the outset of the crisis soon faded away.

This means that trust in government and health care is very important for the outcome of social media discussions. It must be strong enough to survive even tough challenges, such as this 6-year-long struggle. This trust cannot exist only beforehand but must be reinforced during the process.

Conflicts of Interest

None declared.

References

1. European Centre for Disease Prevention and Control. Narcolepsy in association with pandemic influenza vaccination: a multi-country European epidemiological investigation. Stockholm, Sweden: ECDE; 2012 Sep. URL: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/Vaesco%20report%20FINAL%20with%20cover.pdf [accessed 2018-11-02] [WebCite Cache ID 73dQjj2jf]
2. Blomberg K, Carlsson AA, Hagberg L, Jonsson O, Leissner L, Eriksson MH. Quality of life and trust among young people with narcolepsy and their families, after the Pandemrix® vaccination: protocol for a case-control study. BMC Pediatr 2017 Aug 23;17(1):183 [FREE Full text] [doi: 10.1186/s12887-017-0935-4] [Medline: 28835221]

3. Swedish Research Council. The Swedish Research Council's mapping of research relevant to the etiology and treatment of the disease narcolepsy. Stockholm, Sweden: Swedish Research Council; 2012. URL: https://www.vr.se/download/18.2412c5311624176023d255ce/1529480556721/The-Swedish-Research-Councils-Mapping-of-Research.%20Etiology-Treatment-Narcolepsy_VR_2012.pdf [accessed 2018-11-02] [WebCite Cache ID 73dRNZDhU]
4. Abad VC, Guilleminault C. New developments in the management of narcolepsy. Nat Sci Sleep 2017;9:39-57 [FREE Full text] [doi: 10.2147/NSS.S103467] [Medline: 28424564]
5. Hagar C. Crisis informatics: perspectives of trust - is social media a mixed blessing? SLIS Stud Res J 2013;2(2) [FREE Full text]
6. Bennett WL. The personalization of politics. Ann Am Acad Polit Soc Sci 2012 Oct 03;644(1):20-39. [doi: 10.1177/0002716212451428]
7. Kies R. Promises and Limits of Web-Deliberation. New York, NY: Palgrave Macmillan; 2010.
8. Larsson AO. Doing Things in Relation to Machines: Studies on Online Interactivity [doctoral thesis]. Uppsala, Sweden: Uppsala University URL: http://uu.diva-portal.org/smash/get/diva2:512270/FULLTEXT01.pdf [accessed 2019-04-02] [WebCite Cache ID 77KoemtCB]
9. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the human papilloma virus vaccine debate on MySpace blogs. Vaccine 2010 Feb 10;28(6):1535-1540. [doi: 10.1016/j.vaccine.2009.11.060] [Medline: 20003922]
10. Nicholson MS, Leask J. Lessons from an online debate about measles-mumps-rubella (MMR) immunization. Vaccine 2012 May 28;30(25):3806-3812. [doi: 10.1016/j.vaccine.2011.10.072] [Medline: 22063388]
11. Ward JK. Rethinking the antivaccine movement concept: a case study of public criticism of the swine flu vaccine's safety in France. Soc Sci Med 2016 Dec;159:48-57. [doi: 10.1016/j.socscimed.2016.05.003] [Medline: 27173740]
12. Van de Belt TH, Engelen LJLPG, Berben SAA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
13. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
14. Persson I, Granath F, Askling J, Ludvigsson JF, Olsson T, Feltelius N. Risks of neurological and immune-related diseases, including narcolepsy, after vaccination with Pandemrix: a population- and registry-based cohort study with over 2 years of follow-up. J Intern Med 2014 Feb;275(2):172-190 [FREE Full text] [doi: 10.1111/joim.12150] [Medline: 24134219]
15. Persson I, Granath F, Askling J, Ludvigsson JF, Olsson T, Feltelius N. Risks of neurological and immune-related diseases, including narcolepsy, after vaccination with Pandemrix: a population- and registry-based cohort study with over 2 years of follow-up. J Intern Med 2017 Dec;281(1):102-104 [FREE Full text] [doi: 10.1111/joim.12581] [Medline: 27995693]
16. Rieder B. Studying Facebook via data extraction: the netvizz application. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 346-355 URL: http://thepoliticsofsystems.net/permafiles/rieder_websci.pdf
17. World Medical Association. WMA Declaration of Helsinki: ethical principles for medical research involving human subjects. Ferney-Voltaire, France: WMA; 2018 Jul 09. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2018-11-03] [WebCite Cache ID 73eE7lImX]
18. Nielsen J. The 90-9-1 rule for participation inequality in social media and online communities. Fremont, CA: Nielsen Norman Group; 2006. URL: https://www.nngroup.com/articles/participation-inequality/ [accessed 2018-11-02] [WebCite Cache ID 73dNeBvXr]
19. Abramson K, Keefe B, Chou WS. Communicating about cancer through Facebook: a qualitative analysis of a breast cancer awareness page. J Health Commun 2015;20(2):237-243. [doi: 10.1080/10810730.2014.927034] [Medline: 25495200]
20. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12):1-9. [doi: 10.1002/pbc.26660] [Medline: 28618158]
21. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
22. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, et al. A systematic examination of the use of online social networking sites for sexual health promotion. BMC Public Health 2011;11:583 [FREE Full text] [doi: 10.1186/1471-2458-11-583] [Medline: 21777470]
23. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
24. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. “Friending” teens: systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015 Jan 05;17(1):e4 [FREE Full text] [doi: 10.2196/jmir.3692] [Medline: 25560751]

25. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
26. Kazmer MM, Lustria MLA, Cortese J, Burnett G, Kim JH, Ma J, et al. Distributed knowledge in an online patient support community: authority and discovery. J Assoc Inf Sci Technol 2014 Feb 26;65(7):1319-1334. [doi: 10.1002/asi.23064]
27. Rains SA. The implications of stigma and anonymity for self-disclosure in health blogs. Health Commun 2014;29(1):23-31. [doi: 10.1080/10410236.2012.714861] [Medline: 23356432]
28. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
29. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015 Dec;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]

Edited by G Eysenbach; submitted 28.06.18; peer-reviewed by T Ueno, G D’Onofrio; comments to author 07.10.18; revised version received 25.11.18; accepted 29.12.18; published 16.04.19.

Please cite as:
Blomberg K, Eriksson M, Böö R, Grönlund Å
Using a Facebook Forum to Cope With Narcolepsy After Pandemrix Vaccination: Infodemiology Study
J Med Internet Res 2019;21(4):e11419
URL: http://www.jmir.org/2019/4/e11419/
doi:10.2196/11419
PMID:30990457

Original Paper

Technological Innovations in Disease Management: Text Mining US Patent Data From 1995 to 2017

Ming Huang[1], PhD; Maryam Zolnoori[1], PhD; Joyce E Balls-Berry[1], MPE, PhD; Tabetha A Brockman[2,3], MA; Christi A Patten[2,3], PhD; Lixia Yao[1], PhD

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

[2]Center for Clinical and Translational Science, Commuity Engagement Program, Mayo Clinic, Rochester, MN, United States

[3]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Lixia Yao, PhD
Department of Health Sciences Research
Mayo Clinic
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 293 7953
Fax: 1 507 284 1516
Email: Yao.Lixia@mayo.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2019/7/e14678/

Abstract

Background: Patents are important intellectual property protecting technological innovations that inspire efficient research and development in biomedicine. The number of awarded patents serves as an important indicator of economic growth and technological innovation. Researchers have mined patents to characterize the focuses and trends of technological innovations in many fields.

Objective: To expand patent mining to biomedicine and facilitate future resource allocation in biomedical research for the United States, we analyzed US patent documents to determine the focuses and trends of protected technological innovations across the entire disease landscape.

Methods: We analyzed more than 5 million US patent documents between 1995 and 2017, using summary statistics and dynamic topic modeling. More specifically, we investigated the disease coverage and latent topics in patent documents over time. We also incorporated the patent data into the calculation of our recently developed Research Opportunity Index (ROI) and Public Health Index (PHI), to recalibrate the resource allocation in biomedical research.

Results: Our analysis showed that protected technological innovations have been primarily focused on socioeconomically critical diseases such as “other cancers” (malignant neoplasm of head, face, neck, abdomen, pelvis, or limb; disseminated malignant neoplasm; Merkel cell carcinoma; and malignant neoplasm, malignant carcinoid tumors, neuroendocrine tumor, and carcinoma in situ of an unspecified site), diabetes mellitus, and obesity. The United States has significantly improved resource allocation to biomedical research and development over the past 17 years, as illustrated by the decreasing PHI. Diseases with positive ROI, such as ankle and foot fracture, indicate potential research opportunities for the future. Development of novel chemical or biological drugs and electrical devices for diagnosis and disease management is the dominating topic in patented inventions.

Conclusions: This multifaceted analysis of patent documents provides a deep understanding of the focuses and trends of technological innovations in disease management in patents. Our findings offer insights into future research and innovation opportunities and provide actionable information to facilitate policy makers, payers, and investors to make better evidence-based decisions regarding resource allocation in biomedicine.

(J Med Internet Res 2019;21(4):e13316) doi:10.2196/13316

KEYWORDS

patent; technological innovation; disease; research opportunity index; public health index; text mining; topic modeling; dynamic topic model; resource allocation; research priority

Introduction

Patents are an important form of intellectual property that grants inventors monopolies for a limited period of time and provides inventors with a financial incentive for commercialization. Without such financial incentive, private investors in the pharmaceutical and medical device industries may be reluctant to invest in new technologies, which would then slow down the development of new diagnoses and treatments [1]. As patents can promote economically efficient research and development, the number of patents has been used as a proxy for technological innovation and an indicator of economic growth [2]. Patent documents describe the inventor, owner, abstract, claims, and legal status of patented inventions and are publicly available. They have been mined to identify focuses and trends of technological innovations in many areas such as the fisheries sector [3], solar cell industry [4], and drug discovery [5]. A comprehensive survey paper is available to gain a better understanding of the structure of patent documents and the methods for patent document retrieval, classification, and visualization [6].

Existing patent mining in biomedicine mostly focuses on recognizing biomedical entities such as chemical compounds, genes, proteins, cells, tissues, and anatomical parts [7]. For example, Leman et al developed a system of named entity recognition to identify chemical names mentioned in the patents [8]. Fechete et al mined all gene names mentioned in the claim section of diabetic nephropathy-related patents [9]. Grouin applied machine learning approaches to detect pharmacological terms such as target population, organs, symptoms, and treatments [10]. Information mined from patents has been further used to formulate new biomedical hypotheses [9] and discover technological trends about treatment of a specific disease [11]. For instance, Gwak et al identified the trends and leading organizations in wound-healing technology [11] for successful investment strategies and policy making in the future.

One missing aspect in biomedical patent mining is identification of focuses and trends of patented inventions across the entire disease landscape, in order to facilitate evidence-based decision making for future resource allocation in biomedical research. Therefore, in this study, we mined US patent documents from 1995 to 2017 to identify the trends of patent coverage for over 600 diseases and medical conditions. We then incorporated patent coverage for diseases and medical conditions to recalibrate the Research Opportunity Index (ROI) and Public Health Index (PHI) in order to systematically understand resource allocation and research prioritization. ROI and PHI were introduced in our previous work [12] to measure the (im)balance between the health burden associated with a particular disease or medical condition or all diseases and medical conditions as a whole, and the allocated resources. Previously, we used treatment cost as proxies of disease burden and the numbers of scientific publications and clinical trials as indicators of resource allocations. By incorporating patent documents, we considered technological innovation as a driver of resource allocation and research prioritization in biomedicine, which impacts the entire biomedical research ecosystem. Finally, we performed dynamic topic modeling [13] to uncover the latent topics of patented inventions associated with each disease or medical condition and the trend of these topics over time. This study could provide insights into research and development opportunities and offer actionable information for future investment and funding decision making in biomedicine.

Methods

The workflow of this study is illustrated in Figure 1. It includes two phases: (1) data collection and preprocessing and (2) ROI/PHI analysis and topic modeling. Below, we describe each step in more detail.

Patent Data Collection and Filtering

We downloaded approved US patents between 1995 and 2017 from the US Patent and Trademark Office website [14]. The whole dataset included more than 5 million patent documents, whose formats changed three times: Green Book format (1995-2001) [15], Red Book Standard Generalized Markup Language (SGML; 2002-2004) [16], and Red Book XML (2005-2017) [17]. These patent documents were classified into different innovative domains (eg, agriculture, sports, and foodstuffs) based on the United States Patent Classification (USPC) system [18] and the Cooperative Patent Classification (CPC) system [19]. We set the following inclusion criteria to extract patent documents on biomedicine: Publication date should be between January 1, 1995, and December 31, 2017, and the patent document should contain at least one USPC or CPC classification code listed in Multimedia Appendix 1.

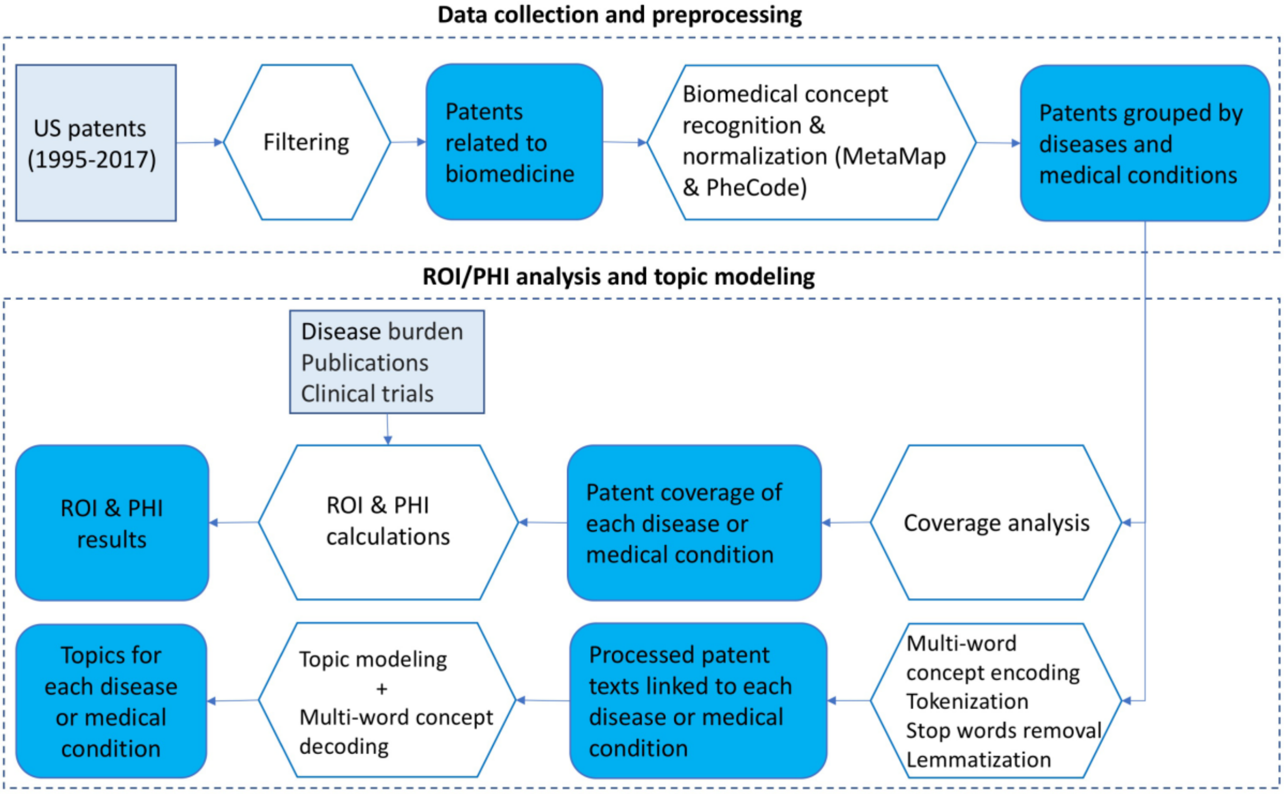
Any documents that did not meet the abovementioned criteria were not considered to be related to biomedicine. We developed a python parser (Multimedia Appendix 2) to parse patent documents in all three formats in order to retrieve information such as patent identification, issue date, title, abstract, claims, and USPC and CPC classification codes. We then filtered for biomedicine-related patents using the compiled lists of USPC and CPC classification codes (Multimedia Appendix 1).

Biomedical Concept Recognition and Normalization

For patent documents related to biomedicine, we used MetaMap, an application developed at the National Library of Medicine [20] to extract and map biomedical concepts (key words or phrases) of 74 semantic types in the sections of title, abstract, and claims to the Unified Medical Language System (UMLS) metathesaurus. The complete list of 74 UMLS semantic types is shown in Multimedia Appendix 1.

UMLS assigns a Concept Unique Identifier (CUI) to each concept and links it to the source thesauri such as the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) [21]. We leveraged ICD-9-CM to identify concepts of diseases and medical conditions in patent documents and mapped ICD-9-CM to PheCode, which represents clinically meaningful phenotypes used by clinicians [22]. To address the issue of concept granularity, we only used the 704 root PheCodes such as diabetes mellitus, influenza, and pain. Thus, patent documents were grouped by diseases and medical conditions before ROI/PHI and topic modeling analysis.

Figure 1. Study workflow of mining US patent data. PHI: Public Health Index; ROI: Research Opportunity Index.



Coverage Analysis

In the second phase, we started by calculating the patent coverage of each disease or medical condition by dividing the number of patent documents mentioning the disease or medical condition by the total number of patent documents mentioning all diseases and medical conditions in each year. This is a preparation step for computing the ROI for each disease or medical condition and PHI for all diseases and medical conditions in each year.

Multiword Concept Encoding, Tokenization, Stop Word Removal, and Lemmatization

To prepare for topic modeling, we encoded multiword biomedical concepts using their corresponding CUIs before tokenizing the patent documents, removing stop words, and lemmatizing words. Multiword concept encoding helps preserve compound concepts such as “type 2 diabetes.” Without multiword concept encoding, such a compound name would be broken down to individual words (unigram) during topic modeling. Tokenization breaks text into smaller meaningful elements such as words, numbers, or punctuation marks. Stop words like “the,” “is,” and “are” are usually filtered out from natural language processing. Lemmatization aims to reduce the morphological variations of words by returning to the base or dictionary form of a word (eg, “walked,” “walking,” and “walks” have the same base form “walk”).

Data on Disease Burden, Publications, and Clinical Trials

The data on disease burden, publications, and clinical trials were collected in the same way as that in our previous work [12]. However, in this study, we included more data from recent years to obtain an updated longitudinal analysis for 17 years (2000-2016). Disease burden was estimated using the total treatment cost in million population each year from OptumLabs Data Warehouse [23]. OptumLabs Data Warehouse is a comprehensive, de-identified administrative claims database for commercially insured and Medicare Advantage enrollees in a large and private US health plan. The diagnosis records were coded by ICD-9-CM before October 2015 and International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) [24] after those claims.

For publications, we downloaded all Medical Subject Heading (MeSH)-indexed abstracts in English from the MEDLINE database (via the PubMed query interface), the biomedical publication database maintained by the United States National Library of Medicine from 2000 to 2016. Subsequently, we used the number of publications annotated by each MeSH disease or medical condition term to approximate the attention it received from the biomedical research community. We also downloaded the aggregated, MeSH-indexed clinical trial database [25]. Similarly, we used the number of clinical trials related to each specific disease or medical condition to approximate feasibility and popularity of carrying out clinical research in each disease or medical condition area. We converted ICD-9-CM, ICD-10-CM, and MeSH codes to the root PheCode before further analysis.

Research Opportunity Index and Public Health Index Calculations

We previously proposed ROI and PHI for quantitatively measuring resource allocation for a particular disease and all the medical conditions as a whole [12]. The calculations of ROI and PHI are flexible and can incorporate many quantitative factors that impact research prioritization and resource allocation in biomedicine. In this work, we updated the ROI and PHI calculations by including more data from recent years and included patent data in addition to disease burden, publications, and clinical trials, since patents are an important form of intellectual property on technological innovations.

ROI and PHI are defined as follows:

where Y_{nd} is the raw measurement n for a disease d. In our model, we used treatment cost per million people as indicators of burden of disease (Y_{bd}) and the number of research publications (Y_{rd}), the number of clinical trials (Y_{cd}), and the number patent documents (Y_{pd}) as an approximation of resources spent in biomedical research and development. Since the raw measurement Y_{nd} is subjective to inflation (eg, increasing treatment cost and number of publications over time) and cannot be compared across different units (eg, treatment cost in dollars vs number of research publications by count), we used the normalized measure X_{nd} instead of Y_{nd} in ROI and PHI calculation. ROI quantifies the imbalance between needs and resource investment on multiple dimensions for each disease or medical condition. PHI describes the overall resource allocation efficiency for all diseases and medical conditions.

Topic Modeling

Topic modeling automatically uncovers the topics or themes in a large collection of documents, in terms of a set of keywords occurring together and most frequently [26-29]. We applied a dynamic topic model (DTM) [13] to learn the topics in patent documents related to specific diseases and medical conditions and the evolution of these topics over years.

DTM is an extension of the static Latent Dirichlet Allocation method [26] for analyzing the temporal changes of topics of a large collection of documents. Static Latent Dirichlet Allocation does not consider the input order of documents in the large collection. It assumes the Dirichlet prior distributions for topic distributions in a document and word distributions over a topic. DTM recognizes the importance of temporality in a large collection of documents and studies the dynamics of topics from time interval $t-1$ to t. It assumes a Gaussian distribution for the prior parameters at t, given their value at $t-1$.

We used the open-source DTM C++ package [30] wrapped in Gensim library [31] to learn the temporality of the topics in patent documents related to a specific disease or medical condition. After multiword concept encoding, tokenization, stop word removal, and lemmatization, each patent document was converted into a vocabulary vector, where the elements were frequency of each lemma (including CUIs) without considering the order of lemma. As such, all the patent documents on the same disease or medical condition were converted into a $D \times V$ matrix, where D stands for count of patent documents and V denotes the size of the entire vocabulary in those patent documents. The $D \times V$ matrix was then chunked into time intervals in the calendar year for dynamic topic modelling.

We calculated topic coherence [32] quantitatively and asked domain experts to qualitatively evaluate the learned topics. More specifically, we evaluated the topic coherence at different topic numbers (ie, 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20) to determine the optimal topic number for three selected diseases: diabetes mellitus, breast cancer, and epilepsy. We found that the optimal topic number was 6 for breast cancer–related patents, 8 for diabetes mellitus–related patents, and 16 for epilepsy-related patents (Multimedia Appendix 1). The average number of topics for the three diseases was 10. Thus, we set the topic number to 10 empirically for the remaining 639 identified diseases and medical conditions.

We also investigated the hyperparameters of α and σ in DTM: α controls the topic distributions over a document, and a smaller α results in fewer topics statistically associated with a document, whereas σ determines how fast the topics evolve over time, and a smaller σ leads to more similar word distributions over a topic over time. In our experiment, we used the default values of 0.01 and 0.005 for α and σ, respectively, as previously suggested [33], as those authors reported that both α and σ did not affect topic distributions and word distributions significantly over time and did not have an effect on topic interpretation by domain experts.

Results

Disease Coverage in Patent Documents During 1995-2017

We collected 5,010,329 patent documents from 1995 to 2017, of which 550,961 (about 11%) were related to biomedicine. Figure 2 shows the percentage of patent documents related to biomedicine and the number of diseases and medical conditions covered in those patent documents in each year. It seemed that the approved US patents on biomedicine fluctuated in the range of 9.6%-14.6% during 1995 and 2017. However, the number of diseases and medical conditions covered in US patents expanded from 502 to 596, suggesting that technology innovations had been focusing on more diseases and medical conditions during the same time.

Figure 2. The percentage of patent documents related to biomedicine (blue dots) and the number of diseases and medical conditions covered in patent documents (orange squares) during 1995-2017.

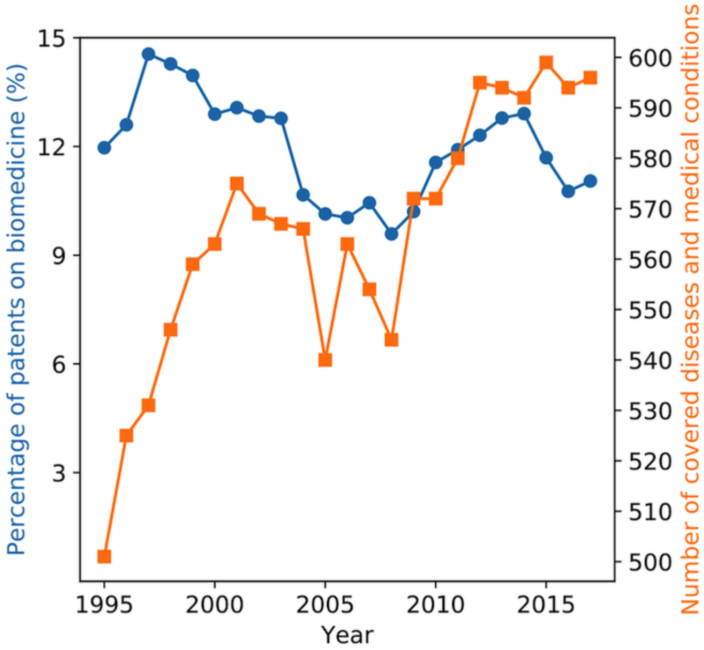
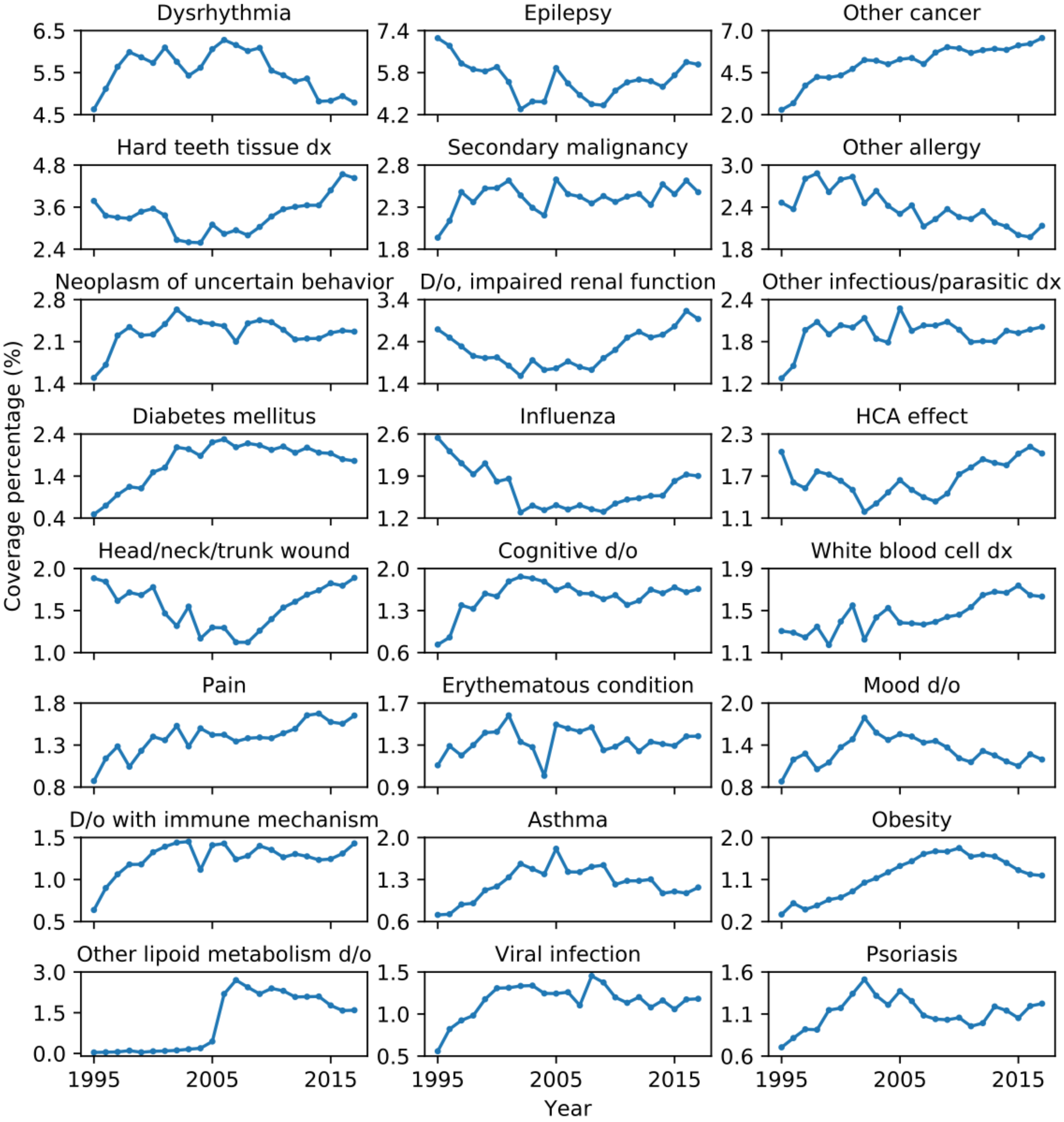


Figure 3 shows the coverage percentages of 24 most mentioned diseases and medical conditions in US patent documents between 1995 and 2017. We found that 16 of them were chronic conditions such as dysrhythmia, "other cancer," diabetes mellitus, and obesity. "Other cancer" refers to malignant neoplasm of head, face, neck, abdomen, pelvis, or limb; disseminated malignant neoplasm; Merkel cell carcinoma; and malignant neoplasm, malignant carcinoid tumors, neuroendocrine tumor, and carcinoma in situ of an unspecified site, according to PheCode [22]. The percentages of patented inventions on "other cancer" are shown to be steadily increasing (from 2.29% in 1995 to 6.56% in 2017). This time period witnessed a steady increase of innovative technologies from functional magnetic resonance imaging to immunotherapy, which improve cancer diagnosis and treatment [34]. With increasing awareness and expenditure on cancer diagnosis and treatment [35], we believe that technological innovations associated with cancer are likely to grow continuously. Similarly, obesity ranked 21st in the most mentioned 24 diseases and medical conditions, and its patent coverage ranged from 0.87% to 1.65% during the study period. Such an increase in patented inventions for obesity aligns well with its high prevalence in the US, the severe implications on people's life and the society, and national strategic plans to deal with the obesity epidemic [36]. Influenza, an acute disease, received decreasing attention in terms of patents before 2000 and increasing attention after 2010, reflecting its high prevalence and substantial burden (eg, large morbidity and mortality) on the health care system and society [37]. The Centers for Disease Control and Prevention reported that influenza caused 9.3-49.0 million illnesses, 140,000-960,000 hospitalizations, and 12,000-79,000 deaths in the United States each year since 2010 [38].

Figure 3. Patent coverage of top 24 mentioned diseases and medical conditions during 1995-2017. Cognitive d/o refers to delirium, dementia, amnesia, and other cognitive disorders. Obesity refers to overweight, obesity, and other hyperalimentation. dx: disease; d/o: disorder; HCA: heat, cold, and air pressure.



Research Opportunity Index and Public Health Index Analysis

Before ROI and PHI analysis, we examined the multicollinearity between the relative number of patent documents and other factors, namely, the relative treatment costs, the relative number of scientific publications, and the relative number of clinic trials used in our previous model [12]. The low variance inflation factors of 1.0-2.6 (Multimedia Appendix 1) indicated a low multicollinearity between the relative number of patent documents and the other factors [39].

We then calculated the ROIs to measure the misalignment between the resources allocated to a disease or medical condition and the burden it imposed between 2000 and 2016, as shown in Figure 4 (see Multimedia Appendix 3 for the raw data). For example, neuroendocrine tumors were overstudied, indicated by negative ROIs from 2000 to 2016. However, their ROI increased from –17.93 (in 2000) to –1.87 (in 2016) over 17 years, implying that resources allocated to it were more aligned with its burden. Further, digging into the dependent variables of ROI, we found that the driving factor for such improvement was that the relative treatment cost (indicator of disease burden) increased much faster than relative publications, relative clinical trials, and relative technological innovations. Similar patterns were observed for most of the overstudied diseases and medical conditions. There were only a few diseases and medical conditions that were more overstudied over time. Injury to nerves not elsewhere classified was such an example. Its ROI scores declined from –6.73 to –9.43 from 2000 to 2016, primarily because its relative treatment cost decreased while it received steady attention in biomedical research (in terms of the relative number of publications) and increasing attention in development (in terms of the relative numbers of clinical trials and patents).

Figure 4 B highlights the understudied diseases and medical conditions with positive ROIs. For example, ankle and foot fracture had steadily declining ROIs (from 12.52 to 6.41) from 2000 to 2016, mostly because the resources allocated to it increased, but its burden slightly decreased over time. More specifically, the relative number of publications increased more than 180 times, the relative number of clinical trials increased over 2 times, and the relative treatment cost decreased by about 20%. The relative numbers of publications, clinical trials, and patents were all disproportionally small, compared to the burden of ankle and foot fracture. In contrast, intracranial hemorrhage (injury) had gradually increasing ROIs (from 6.31 in 2000 to 10.19 in 2016) and was becoming more understudied due to increase in the relative treatment cost and decline in the relative number of publications. There were also diseases and medical conditions that showed fluctuating ROIs over time. For instance, the ROIs for contact dermatitis increased from 6.90 to 8.20 during 2000-2014 and decreased to 5.63 in 2016. Our calculation showed that its relative treatment cost declined steadily by a factor of 1.6 from 2000 to 2016, but its relative number of publications and patents fluctuated dramatically during the same period.

The overall alignment between disease burden and allocated resources for all the diseases and medical conditions measured by PHI from 2000 to 2016 is shown in Figure 5. It demonstrated a clear shrinking pattern over the 17-year period, with small increases in 2015 and 2016. As the smaller PHI indicates better alignment between research and development and the distributions of needs across all diseases and medical conditions, the results suggest that the resource allocation for all the diseases and medical conditions as a whole had improved significantly over time.

Topic Modeling

Using a dynamic topic modeling technique, we identified the latent topics and their changing patterns over time for patent documents related to various diseases and medical conditions. In Figure 6, we highlight three meaningful topics identified from patent documents related to diabetes mellitus, breast cancer, and epilepsy, together with their changing patterns from 1995 to 2017. The results for the other diseases and medical conditions are provided in Multimedia Appendix 3.

Figure 4. Research Opportunity Index visualization for overstudied (A) and understudied (B) diseases and medical conditions using co-centric circles. A different color of a circle corresponds to a different year, illustrated by the legend on the left. (A) The size of a circle enlarges as the negative Research Opportunity Index of the corresponding disease or medical condition decreases. In other words, the bigger the circle is, the more overstudied is the disease or medical condition. (B) The size of each circle enlarges as the positive Research Opportunity Index of the disease or medical condition increases. In other words, the bigger the circle is, the more understudied is the disease or medical condition, indicating a future research opportunity. Contact dermatitis refers to contact dermatitis and other eczema due to plants except food. dx: disease; NEC: not elsewhere classified; CRP: C-reactive protein; SIRS: systemic inflammatory response syndrome; IVS: intracranial venous sinuses; ICH: intracranial hemorrhage.

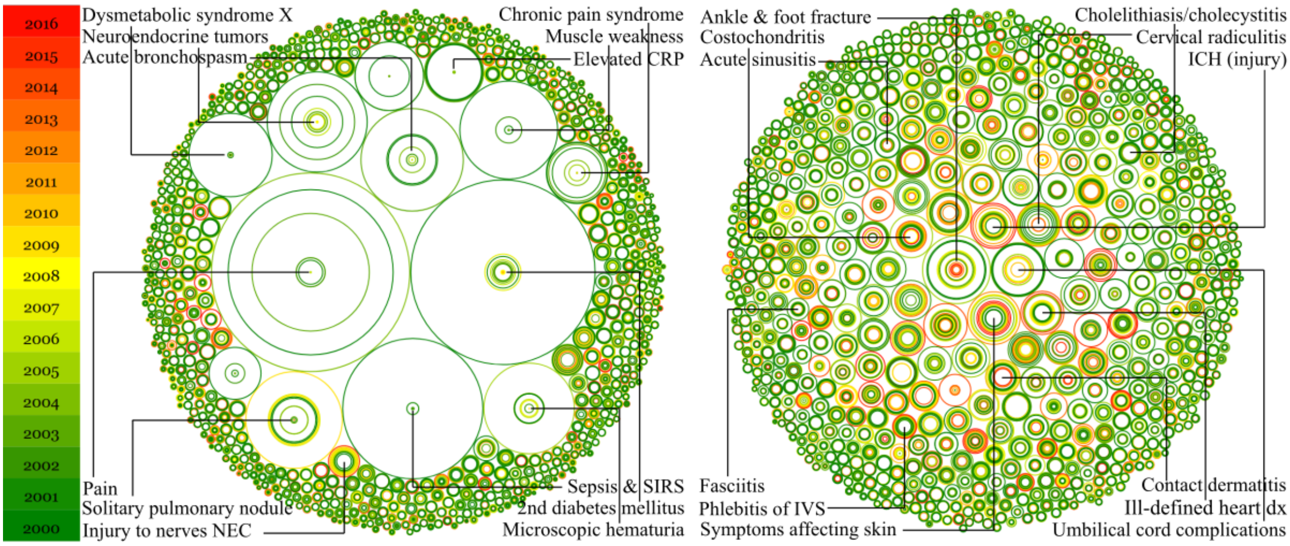


Figure 5. Public Health Index (PHI) during 2000-2016.

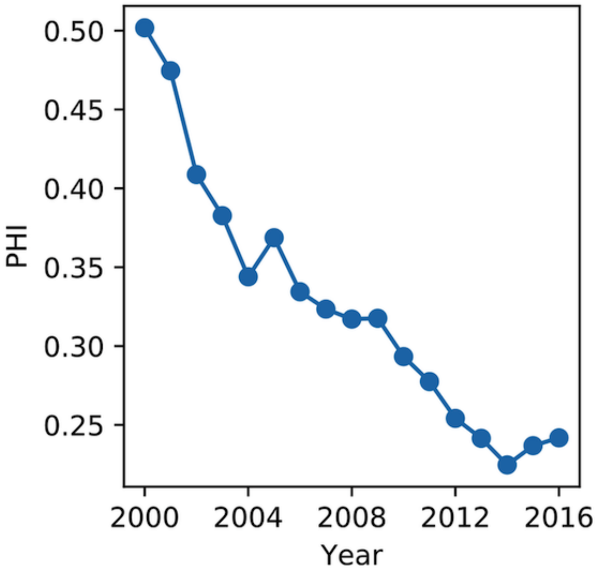
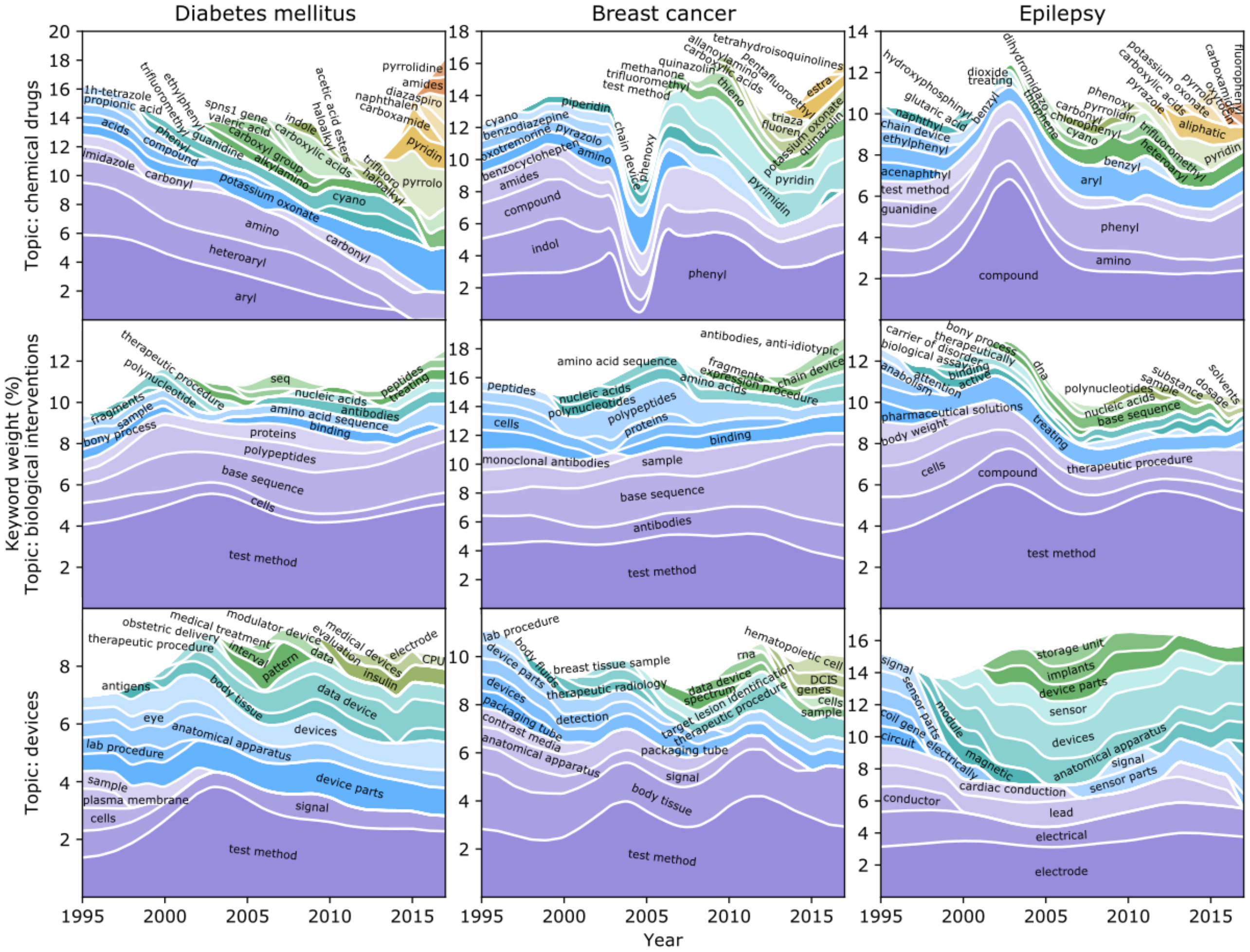


Figure 6. Three meaningful topics identified from patent documents related to diabetes mellitus, breast cancer, and epilepsy, and their changing patterns from 1995 to 2017.



Diabetes Mellitus

With a prevalence of 9.4% among the US population and a financial burden of $245 billion, diabetes mellitus was the seventh leading cause of death in the US in 2015 [40]. Both public and private sectors in the health care industry have invested heavily in developing new methods of diagnosis and treatment for diabetes mellitus. Our dynamic topic modeling analysis showed that most of the patent activities revolved around the development of new drugs (both chemical compounds and biological therapeutics) and new devices to enhance glucose monitoring and diabetes management. More specifically, keywords including “aryl,” “heteroaryl,” and “phenyl” indicate that chemical drugs are a main focus of those patent documents [41]. The top 10 topic keywords also evolved over time; for example, “propionic acid” only appeared during 1995-2000 [42], “valeric acid” appeared during 2004-2006 [43], “indole” appeared during 2008-2009 [43,44], “naphthalen” appeared in 2015 [45], and “diazaspiro” appeared during 2016-2017 [46]. These keywords highlighted the significance of different chemical compounds or functional groups in the development of antidiabetic drugs over years. The topic keywords “base sequence,” “polypeptides,” “antibodies,” and “therapeutic procedure” revealed that biological interventions, gene therapy [47], peptide therapy [48], and immunotherapy [49] were other focuses of the patented innovations. After tracing back to the original patent documents, we found that the identified keywords “signal,” “device parts,” and “devices” were related to devices for improving monitoring [50] and management [51] of glucose level.

Breast Cancer

Breast cancer is the most common cancer worldwide [52] and has gained a lot of attention for innovations on diagnosis and therapeutic treatment. Its 5-year relative survival rate increased from 75% in 1975-1977 to 91% in 2006-2012 [53]. The focus of patented inventions related to breast cancer was primarily on the development of novel pharmaceutical drugs, biological products, and devices for diagnosis and treatment. Chemical agents or functional groups such as “phenyl” (1995-2017) [54], “pyrazolo” (1995-2003) [55], “thieno” (2008-2013) [56], and “tetrahydroisoquinolines” (2017) [57] were most mentioned in the top 10 topic words and were reported to have antitumor effects in different years. In addition, “antibodies,” “binding,” “base sequence,” “polypeptides,” and “amino acid sequence” were the most frequent keywords of breast cancer mentioned in patents related to antibody therapy [58] and genetic diagnosis and treatment [59,60]. The keywords on antibody such as “monoclonal antibodies” (1995-2003) and “antibodies, anti-idiotypic” (2014-2017) further reflect the development of patented inventions on antibody therapy for breast cancer [61,62]. The topic words “body tissue,” “signal,” “contrast media,” “devices,” “detection,” “device parts,” “therapeutic radiology,” and “therapeutic procedure” are associated with devices for the detection and treatment for breast cancer [52,63].

Epilepsy

Epilepsy is one of the most common neurologic diseases involving about 50 million people worldwide and 3 million people in the United States [64]. It is a chronic disorder that usually consists of unpredictable recurrent seizures with substantial impacts on patients’ mental and physical functioning. Important topics in epilepsy-related patent documents were relevant therapeutic drugs, biological interventions, and electrical devices for epilepsy diagnosis and treatment. During the study period, the patent topic on chemicals involved different keywords such as “phenyl” (1995-2017) [65,66], “naphthyl” (1995-1999) [67], “thiophene” (2004-2005) [68], “pyrrolidin” (2010) [69], and “carboxamide” (2016-2017) [70]. These keywords suggested that these chemical compounds or functional groups were the leading efforts in antiepileptic drug development over years. The topic words “cells,” “therapeutic procedure,” “biological assay,” “base sequence,” “nucleic acids,” and “polynucleotides” disclose patent inventions on biological interventions and gene therapy [71,72]. The keywords “electrode,” “signal,” “devices,” “sensor,” and “implants” suggested that electrical devices for epilepsy diagnosis and intervention were also the focus of patented inventions [73-75].

Discussion

In this study, we identified 550,961 biomedicine-related patent documents; calculated patent coverage, ROI, and PHI; and performed topic modeling analysis for more than 600 diseases and medical conditions from about two decades. We found that technological innovations reached an increasing number of diseases and medical conditions from 1995 to 2017. The innovation hotspots were around common chronic conditions including “other cancer,” diabetes mellitus, and obesity, which bore significant socioeconomic burden [76]. Technological inventions related to acute conditions, such as influenza, also attained substantial attention due to their high morbidity and mortality [37]. Unfortunately, patents, as a financial incentive to intellectual properties, have not penetrated into many rare diseases yet.

Calculation of PHI from 2000 to 2016 clearly demonstrated that the resource allocation for all the diseases and medical conditions as a whole had improved significantly over time. This is consistent with our previous findings [12], suggesting that the overall resource allocation in biomedical research and development has been improving significantly in the United States, possibly due to more available quantitative data from epidemiology studies and improved transparency in biomedical research and development. Disease-specific ROI tells us whether the resources allocated to a disease align with its burden imposed on the society. The skewness between allocated resources and disease burden measured by treatment cost improved for most overstudied diseases and medical conditions from 2000 to 2016, which possibly contributed to the overall improvement of PHI for all diseases and medical conditions. A few diseases and medical conditions became more overstudied, which demonstrated substantial “inertia” in allocated resources including publications, clinical trials, and patents and their disconnection with disease burden from previous years. One possible explanation is that there was no feedback mechanism to realign resource allocation with disease burden or that the relationship is complex and mediated by other observed variables. For example, researchers’ attention is influenced by exposure to health problems that appear in their local hospitals and clinics, and they have to maintain a relatively stable disease focus for funding and publication purposes in their research career. Negative feedback between patents, clinical trials, scientific studies, and disease burden exists, but operates on a longer timescale than we were able to observe in this study. Diseases such as ankle and foot fracture and intracranial hemorrhage (injury) received positive ROI and thus form a niche for future research and development opportunity.

Additional topic modeling expectedly showed that technological innovations largely focused on developing new diagnosis and treatment for most common chronic and acute diseases and medical conditions, which is in line with several qualitative studies from manual analysis of patent documents [77-79]. The evolution of topic keywords reflects technological development (eg, chemical drugs) on diseases and medical conditions over years.

This study has several limitations. First, disease is not a properly defined concept. Here, we used “diseases and medical conditions” to refer to diseases, syndromes, disorders, symptoms, and abnormalities, as long as they are treated by providers and included in the PheCode taxonomy. No existing medical taxonomy is able to address the issues of granularity, disease comorbidity, and association and the distinction between diseases and symptoms perfectly. However, such imperfection does not revoke the significance of this work, because we are addressing a macroeconomic problem of resource allocation and optimization in biomedicine. Second, there are cases that name variations of biomedical concepts were not listed in the UMLS metathesaurus or MetaMap failed to recognize a disease name and map it correctly to the UMLS metathesaurus [80]. Third, we grouped the recognized CUI for diseases and medical conditions to the root PheCode via ICD-9-CM [21] using mapping tables provided by UMLS and the PheCode team. The percentage of definitive mappings (eg, one to one and multiple to one) from CUI to PheCode is 98.4%, which suggests that the upper bound of error caused by ambiguous mappings (eg, one to multiple or multiple to multiple mappings) might be 1.6%. In addition, in ROI and PHI analysis, we converted ICD-10-CM used in claims database to the root PheCode using ICD-9-CM as a middle layer, as claims database switched from ICD-9-CM to ICD-10-CM in 2015 for coding diseases and medical conditions. We also mapped MeSH terms used by publications and clinical trials to the root PheCode. Imperfectness in the mapping between disease taxonomies could lead to result inaccuracy and interpretation difficulty. Fourth, we used the treatment costs estimated from a large claims database to approximate burdens of diseases and medical conditions when computing the ROI and PHI. We acknowledge that such approximation is far from being perfect because information about uninsured people and uncovered ailments were missing, and human suffering from each disease or medical condition cannot be fully measured by treatment costs. Our choice was a compromise, as there are no objective and comparable measures of disease burden for the entire disease landscape. Finally, state-of-the-art DTM exploits statistical inference built on term frequency when identifying the latent patterns. Therefore, high-frequency terms are likely to dominate the identified topics, which limits our capability to identify rare, yet meaningful, topics from patent documents. Furthermore, tuning the hypermeters in DTM can be more art than science.

Acknowledgments

Funding for this study was provided by Mayo Clinic Center for Clinical and Translational Science (UL1TR002377), the National Institutes of Health/National Center for Advancing Translational Sciences (NIH/NCATS), and the National Library of Medicine (5K01LM012102).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The US patent classification systems and biomedicine-relevant codes (Table S1); UMLS semantic types on biomedicine (Table S2); Summary statistics of US patent data during 1995-2017 (Table S3); Multicollinearity between the relative number of patents and other factors including the relative treatment cost, the relative number of publications and the relative number of clinical trials (Table S4); Coherence scores of learned topics over different topic number for diabetes mellitus, breast cancer, and epilepsy (Figure S1).

[PDF File (Adobe PDF File), 173KB - jmir_v21i4e13316_app1.pdf]

Multimedia Appendix 2

Python script to parse patent documents in 3 different formats.

[TXT File, 21KB - jmir_v21i4e13316_app2.txt]

Multimedia Appendix 3

Patent coverage (Table S5), ROIs (Table S6), PHIs (Table S7), and topics (Tables S8 and S9) for over 600 diseases and medical conditions.

[XLSX File (Microsoft Excel File), 7MB - jmir_v21i4e13316_app3.xlsx]

References

1. Cockburn I, Long G. The importance of patents to innovation: updated cross-industry comparisons with biopharmaceuticals. Expert Opin Ther Pat 2015 Jul;25(7):739-742. [doi: 10.1517/13543776.2015.1040762] [Medline: 25927945]
2. Raghupathi V, Raghupathi W. Innovation at country-level: association between economic development and patents. J Innov Entrep 2017 Feb 27;6(1):4. [doi: 10.1186/s13731-017-0065-0]
3. Vincent C, Singh V, Chakraborty K, Gopalakrishnan A. Patent data mining in fisheries sector: An analysis using Questel-Orbit and Espacenet. World Patent Information 2017 Dec 27;51(1):22-30. [doi: 10.1016/j.wpi.2017.11.004]
4. Tseng F, Hsieh C, Peng Y, Chu Y. Using patent data to analyze trends and the technological strategies of the amorphous silicon thin-film solar cell industry. Technological Forecasting and Social Change 2011 Feb;78(2):332-345. [doi: 10.1016/j.techfore.2010.10.010]
5. Grandjean N, Charpiot B, Pena C, Peitsch M. Competitive intelligence and patent analysis in drug discovery. Drug Discov Today Technol 2005;2(3):211-215 [FREE Full text] [doi: 10.1016/j.ddtec.2005.08.007] [Medline: 24981938]
6. Zhang L, Li L, Li T. Patent Mining. SIGKDD Explor Newsl 2015 May 21;16(2):1-19. [doi: 10.1145/2783702.2783704]
7. Rodriguez-Esteban R, Bundschus M. Text mining patents for biomedical knowledge. Drug Discov Today 2016 Dec;21(6):997-1002. [doi: 10.1016/j.drudis.2016.05.002] [Medline: 27179985]
8. Leaman R, Wei C, Zou C, Lu Z. Mining chemical patents with an ensemble of open systems. Database (Oxford) 2016;2016:baw065 [FREE Full text] [doi: 10.1093/database/baw065] [Medline: 27173521]
9. Fechete R, Heinzel A, Perco P, Mönks K, Söllner J, Stelzer G, et al. Mapping of molecular pathways, biomarkers and drug targets for diabetic nephropathy. Proteomics Clin Appl 2011 Jun;5(5-6):354-366. [doi: 10.1002/prca.201000136] [Medline: 21491608]
10. Grouin C. Biomedical entity extraction using machine-learning based approaches. 2014 Presented at: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014); May 2014; Reykjavik, Iceland p. 2518-2523 URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/236_Paper.pdf
11. Gwak J, Sohn S. Identifying the trends in wound-healing patents for successful investment strategies. PLoS One 2017;12(3):e0174203 [FREE Full text] [doi: 10.1371/journal.pone.0174203] [Medline: 28306732]
12. Yao L, Li Y, Ghosh S, Evans J, Rzhetsky A. Health ROI as a measure of misalignment of biomedical needs and resources. Nat Biotechnol 2015 Dec;33(8):807-811 [FREE Full text] [doi: 10.1038/nbt.3276] [Medline: 26252133]
13. Blei D, Lafferty J. Dynamic topic models. 2006 Presented at: Proceedings of the 23rd international conference on Machine learning 2006; June 2006; Pittsburgh, Pennsylvania, USA p. 113-120 URL: https://dl.acm.org/citation.cfm?id=1143859 [doi: 10.1145/1143844.1143859]
14. USPTO. Bulk Data Storage System (BDSS) Version 1.1.0 URL: https://bulkdata.uspto.gov/ [accessed 2019-04-24] [WebCite Cache ID 77sQRJXjo]
15. USPTO. United States Patent and Trademark Office Patent Grant Full Text Data/APS URL: https://bulkdata.uspto.gov/data/patent/grant/redbook/fulltext/1976/PatentFullTextAPSDoc_GreenBook.pdf [accessed 2019-04-24]
16. USPTO. 2001. Grant Red Book: Specification for SGML Markup of United States Patent Grant Publications V2.4 URL: https://www.uspto.gov/sites/default/files/products/PatentGrantSGMLv24-Documentation.pdf [accessed 2019-04-24] [WebCite Cache ID 77sRs6JD4]
17. USPTO. 2004. XML Resources URL: https://www.uspto.gov/learning-and-resources/xml-resources [accessed 2019-04-23] [WebCite Cache ID 77sRWknY2]
18. USPTO. 2012. Overview of the U.S. Patent Classification System (USPC) URL: https://www.uspto.gov/sites/default/files/patents/resources/classification/overview.pdf [accessed 2019-04-24] [WebCite Cache ID 77sRaBy9W]
19. European Patent Office and United States Patent and Trademark Office. 2013. Cooperative Patent Classification System URL: https://www.cooperativepatentclassification.org/index.html [accessed 2019-04-24] [WebCite Cache ID 77sRjskPl]
20. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
21. Centers for Disease Control and Prevention. 1996. International classification of diseases, ninth revision, clinical modification (ICD-9-CM) URL: https://www.cdc.gov/nchs/icd/icd9cm.htm [accessed 2019-04-24] [WebCite Cache ID 77sSSUack]
22. Wei W, Bastarache L, Carroll R, Marlo J, Osterman T, Gamazon E, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One 2017;12(7):e0175508 [FREE Full text] [doi: 10.1371/journal.pone.0175508] [Medline: 28686612]
23. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Bleicher PD, Crown WH. Optum Labs: building a novel node in the learning health care system. Health Aff (Millwood) 2014 Jul;33(7):1187-1194 [FREE Full text] [doi: 10.1377/hlthaff.2014.0038] [Medline: 25006145]
24. CDC. ICD-10-CM Official Guidelines for Coding and Reporting FY 2019 URL: https://www.cdc.gov/nchs/icd/data/10cmguidelines-FY2019-final.pdf [accessed 2019-04-24] [WebCite Cache ID 77se7uYzE]
25. U.S. National Library of Medicine: ClinicalTrials.gov. URL: https://clinicaltrials.gov/ [accessed 2019-04-24] [WebCite Cache ID 77sT9Fokq]
26. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993-1022.
27. Huang M, ElTayeby O, Zolnoori M, Yao L. Public Opinions Toward Diseases: Infodemiological Study on News Media Data. J Med Internet Res 2018 Dec 08;20(5):e10047 [FREE Full text] [doi: 10.2196/10047] [Medline: 29739741]

28. Bian J, Zhao Y, Salloum R, Guo Y, Wang M, Prosperi M, et al. Using Social Media Data to Understand the Impact of Promotional Information on Laypeople's Discussions: A Case Study of Lynch Syndrome. J Med Internet Res 2017 Dec 13;19(12):e414 [FREE Full text] [doi: 10.2196/jmir.9266] [Medline: 29237586]
29. Chen A, Zhu S, Conway M. What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques. J Med Internet Res 2015 Sep 29;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]
30. Blei D. GitHub. Dynamic Topic Models and the Document Influence Model URL: https://github.com/blei-lab/dtm [accessed 2019-04-24] [WebCite Cache ID 77sTb6G9h]
31. Rehurek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. : University of Malta; 2010 Presented at: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks; May 22, 2010; Valletta, Malta, ELRA p. 46-50 URL: http://is.muni.cz/publication/884893/en
32. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. 2015 Presented at: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; February 2015; Shanghai, China p. 399-408 URL: https://dl.acm.org/citation.cfm?id=2685324
33. Lee M, Liu Z, Huang R, Tong W. Application of dynamic topic models to toxicogenomics data. BMC Bioinformatics 2016 Oct 06;17(Suppl 13):368 [FREE Full text] [doi: 10.1186/s12859-016-1225-0] [Medline: 27766956]
34. Lee H, Chen Y. Image based computer aided diagnosis system for cancer detection. Expert Systems with Applications 2015 Jul;42(12):5356-5365. [doi: 10.1016/j.eswa.2015.02.005]
35. National Cancer Institute. 2018 Apr 27. Cancer Statistics URL: https://www.cancer.gov/about-cancer/understanding/statistics [accessed 2019-04-24] [WebCite Cache ID 77sXbfhOq]
36. Dietz W. The response of the US Centers for Disease Control and Prevention to the obesity epidemic. Annu Rev Public Health 2015 Mar 18;36:575-596. [doi: 10.1146/annurev-publhealth-031914-122415] [Medline: 25581155]
37. Rolfes M, Foppa I, Garg S, Flannery B, Brammer L, Singleton J, et al. Annual estimates of the burden of seasonal influenza in the United States: A tool for strengthening influenza surveillance and preparedness. Influenza Other Respir Viruses 2018 Dec;12(1):132-137 [FREE Full text] [doi: 10.1111/irv.12486] [Medline: 29446233]
38. Centers for Disease Control Prevention. 2018. Disease Burden of Influenza URL: https://www.cdc.gov/flu/about/burden/index.html [accessed 2019-04-24] [WebCite Cache ID 77sXuqUQy]
39. Gareth J, Daniela W, Trevor H, Robert T. An introduction to statistical learning: with applications in R. New York: Springer; 2019.
40. Centers for Disease Control Prevention. 2017 Jun 18. New CDC report: More than 100 million Americans have diabetes or prediabetes URL: https://www.cdc.gov/media/releases/2017/p0718-diabetes-report.html [accessed 2019-04-24] [WebCite Cache ID 77sYsEZsP]
41. Kumar A, Chawla A, Jain S, Kumar P, Kumar S. 3-Aryl-2-{4-[4-(2,4-dioxothiazolidin-5-ylmethyl)phenoxy]-phenyl}-acrylic acid alkyl ester: synthesis and antihyperglycemic evaluation. Med Chem Res 2010 Jun 5;20(6):678-686. [doi: 10.1007/s00044-010-9369-3]
42. Connell R. Glucagon antagonists for the treatment of Type 2 diabetes. Expert Opinion on Therapeutic Patents 2005 Feb 25;9(6):701-709. [doi: 10.1517/13543776.9.6.701]
43. Bhaskaran S, Mohan V. United States Patent. 2006. Synergistic composition for the treatment of diabetes mellitus URL: https://patents.google.com/patent/US7141254B2/en [accessed 2019-04-25] [WebCite Cache ID 77tszYx7K]
44. Xiong X, Pirrung M. Modular synthesis of candidate indole-based insulin mimics by Claisen rearrangement. Org Lett 2008 Mar 20;10(6):1151-1154. [doi: 10.1021/ol800058d] [Medline: 18303898]
45. Wu J, Lu C, Li X, Fang H, Wan W, Yang Q, et al. Synthesis and Biological Evaluation of Novel Gigantol Derivatives as Potential Agents in Prevention of Diabetic Cataract. PLoS One 2015;10(10):e0141092 [FREE Full text] [doi: 10.1371/journal.pone.0141092] [Medline: 26517726]
46. Hirose H, Yamasaki T, Ogino M, Mizojiri R, Tamura-Okano Y, Yashiro H, et al. Discovery of novel 5-oxa-2,6-diazaspiro[3.4]oct-6-ene derivatives as potent, selective, and orally available somatostatin receptor subtype 5 (SSTR5) antagonists for treatment of type 2 diabetes mellitus. Bioorg Med Chem 2017 Dec 01;25(15):4175-4193. [doi: 10.1016/j.bmc.2017.06.007] [Medline: 28642028]
47. Oh S, Lee M, Ko K, Choi S, Kim S. GLP-1 gene delivery for the treatment of type 2 diabetes. Molecular Therapy 2003 Apr;7(4):478-483. [doi: 10.1016/S1525-0016(03)00036-4]
48. Rosenberg L. United States Patent. 2016. Modified Ingap Peptides For Treating Diabetes URL: https://patents.google.com/patent/US20160002310A1/en?oq=US20160002310A1 [accessed 2019-04-25] [WebCite Cache ID 77ttEz6IH]
49. Khan NA, Benner R. United States Patent. 2009. Treatment of type I diabetes URL: https://patents.google.com/patent/US7517529B2/en?oq=US7517529B2 [accessed 2019-04-25] [WebCite Cache ID 77ttdjcLL]
50. Pinsker JE. United States Patent. 2013. Diabetes Monitoring Using Smart Device URL: https://patents.google.com/patent/US20130332196A1/en?oq=US20130332196A1 [accessed 2019-04-25] [WebCite Cache ID 77ttrK4aR]
51. Reinke RE, Price JF, Galley PJ. United States Patent. 2011. Diabetes health management systems and methods URL: https://patents.google.com/patent/US20110124996A1/en?oq=US20110124996A1 [accessed 2019-04-25] [WebCite Cache ID 77ttzadZB]

52. Nounou M, ElAmrawy F, Ahmed N, Abdelraouf K, Goda S, Syed-Sha-Qhattal H. Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies. Breast Cancer (Auckl) 2015;9(Suppl 2):17-34 [FREE Full text] [doi: 10.4137/BCBCR.S29420] [Medline: 26462242]
53. American Cancer Society. 2017. Cancer Facts & Figures 2017 URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf [accessed 2019-04-24] [WebCite Cache ID 77sateNzD]
54. Ramana M, Lokhande R, Bhar S, Ranade P, Mehta A, Gadre G. In Silico Design, Synthesis and Bioactivity of N-(2,4-Dinitrophenyl)-3-oxo-3-phenyl-N-(aryl) Phenyl Propanamide Derivatives as Breast Cancer Inhibitors. Curr Comput Aided Drug Des 2017;13(2):112-126. [doi: 10.2174/1573409912666161223160217] [Medline: 28019636]
55. Dias L, Freitas A, Barreiro E, Goins D, Nanayakkara D, McChesney J, et al. Synthesis and biological activity of new potential antimalarial: 1H-pyrazolo[3,4-b]pyridine derivatives. Boll Chim Farm 2000;139(1):14-20. [Medline: 10829547]
56. Kandeel M, Abdelhameid MK, Eman K, Labib MB. Synthesis of some novel thieno[3,2-d]pyrimidines as potential cytotoxic small molecules against breast cancer. Chem Pharm Bull (Tokyo) 2013;61(6):637-647 [FREE Full text] [Medline: 23538397]
57. Zhu P, Ye W, Li J, Zhang Y, Huang W, Cheng M, et al. Design, synthesis, and biological evaluation of novel tetrahydroisoquinoline derivatives as potential antitumor candidate. Chem Biol Drug Des 2017 Dec;89(3):443-455. [doi: 10.1111/cbdd.12873] [Medline: 27717183]
58. Willems A, Gauger K, Henrichs C, Harbeck N. Antibody therapy for breast cancer. Anticancer Res 2005;25(3A):1483-1489 [FREE Full text] [Medline: 16033049]
59. Gomis CR, Tarragona SM, Arnal EA, Pavlovic M. United States Patent. 2014. Method for the diagnosis, prognosis and treatment of breast cancer metastasis URL: https://patents.google.com/patent/US10047398B2/en?oq=US10047398B2 [accessed 2019-04-25] [WebCite Cache ID 77tuE7jA9]
60. Chen HM. United States Patent. 2002. Genes expressed in breast cancer URL: https://patents.google.com/patent/US20020156263A1/en?oq=US20020156263A1 [accessed 2019-04-25] [WebCite Cache ID 77tuMetQa]
61. Green M, Murray J, Hortobagyi G. Monoclonal antibody therapy for solid tumors. Cancer Treat Rev 2000 Aug;26(4):269-286. [doi: 10.1053/ctrv.2000.0176] [Medline: 10913382]
62. Ladjemi M. Anti-idiotypic antibodies as cancer vaccines: achievements and future improvements. Front Oncol 2012;2:158 [FREE Full text] [doi: 10.3389/fonc.2012.00158] [Medline: 23133825]
63. Wörmann B. Breast cancer: basics, screening, diagnostics and treatment. Med Monatsschr Pharm 2017 Feb;40(2):55-64. [Medline: 29952495]
64. Goldenberg M. Overview of drugs used for epilepsy and seizures: etiology, diagnosis, and treatment. P T 2010 Jul;35(7):392-415 [FREE Full text] [Medline: 20689626]
65. Sandoval G, Toledo S, Garcia M. United States Patent. 1995. Phenyl alcohol amides having anticonvulsant activity URL: https://patents.google.com/patent/US5463125A/en?oq=US5463125A [accessed 2019-04-25]
66. Choi Y. United States Patent. 2017. Phenyl carbamate compounds for use in preventing or treating epilepsy URL: https://patents.google.com/patent/US9624164B2/en?oq=US9624164B2 [accessed 2019-04-25] [WebCite Cache ID 77tv2mazT]
67. Grigor'ev V, Kalashnikov V, Kalashnikova I, Nemanova V, Tkachenko S, Bachurin S. Synthesis and study of the anticonvulsant action of n-naphthyl-(4-pyridinyl)methylamines and their analogs. Pharm Chem J 1999 Aug;33(8):403-405. [doi: 10.1007/BF02510086]
68. Rückle T, Biamonte M, Grippi-Vallotton T, Arkinstall S, Cambet Y, Camps M, et al. Design, synthesis, and biological activity of novel, potent, and selective (benzoylaminomethyl)thiophene sulfonamide inhibitors of c-Jun-N-terminal kinase. J Med Chem 2004 Dec 30;47(27):6921-6934. [doi: 10.1021/jm031112e] [Medline: 15615541]
69. Li X, Zhu C, Li C, Wu K, Huang D, Huang L. Synthesis of N-substituted Clausenamide analogues. Eur J Med Chem 2010 Nov;45(11):5531-5538. [doi: 10.1016/j.ejmech.2010.08.041] [Medline: 20864223]
70. Ahsan M. Anticonvulsant activity and neuroprotection assay of 3-substituted-N-aryl-6,7-dimethoxy-3a,4-dihydro-3H-indeno[1,2-c]pyrazole-2-carboxamide analogues. Arabian Journal of Chemistry 2017 May;10:S2762-S2766. [doi: 10.1016/j.arabjc.2013.10.023]
71. Simonato M. Gene therapy for epilepsy. Epilepsy Behav 2014 Sep;38:125-130. [doi: 10.1016/j.yebeh.2013.09.013] [Medline: 24100249]
72. Beyenburg S, Watzka M, Blümcke I, Schramm J, Bidlingmaier F, Elger C, et al. Expression of mRNAs encoding for 17beta-hydroxisteroid dehydrogenase isozymes 1, 2, 3 and 4 in epileptic human hippocampus. Epilepsy Res 2000 Aug;41(1):83-91. [Medline: 10924871]
73. Kramer U, Shaham A, Shpitalnik S, Weissman N, Goren Y, Kartoun U. United States Patent. 2012. Device and method for detecting an epileptic event URL: https://patents.google.com/patent/US8109891B2/en?oq=US8109891B2 [accessed 2019-04-25] [WebCite Cache ID 77tvDFiaF]
74. Whitehurst TK, McGivern JP, Kuzma JA. United States Patent. 2004. Fully implantable miniature neurostimulator for stimulation as a therapy for epilepsy URL: https://patents.google.com/patent/US6788975B1/en?oq=US6788975B1 [accessed 2019-04-25] [WebCite Cache ID 77tvMe3B8]
75. Pless B. United States Patent. 2007. Electrical stimulation strategies to reduce the incidence of seizures URL: https://patents.google.com/patent/US7174213B2/en?oq=US7174213B2 [accessed 2019-04-25] [WebCite Cache ID 77tvV3f4W]

76. Bernell S, Howard S. Use Your Words Carefully: What Is a Chronic Disease? Front Public Health 2016;4:159 [FREE Full text] [doi: 10.3389/fpubh.2016.00159] [Medline: 27532034]
77. Boehm M, Crawford M, Moscovitz J, Carpino P. Diabetes area patent participation analysis - part II: years 2011-2016. Expert Opin Ther Pat 2018 Dec;28(2):111-122. [doi: 10.1080/13543776.2018.1406477] [Medline: 29140125]
78. Ortiz R, Melguizo C, Prados J, Álvarez PJ, Caba O, Rodríguez-Serrano F, et al. New gene therapy strategies for cancer treatment: a review of recent patents. Recent Pat Anticancer Drug Discov 2012 Sep;7(3):297-312. [Medline: 22339358]
79. Ramos M, Boulaiz H, Griñan-Lison C, Marchal J, Vicente F. What's new in treatment of pancreatic cancer: a patent review (2010-2017). Expert Opin Ther Pat 2017 Nov;27(11):1251-1266. [doi: 10.1080/13543776.2017.1349106] [Medline: 28665163]
80. Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Res 2012 Jan;40(Database issue):D940-D946 [FREE Full text] [doi: 10.1093/nar/gkr972] [Medline: 22080554]

Abbreviations

CPC: Cooperative Patent Classification
CUI: Concept Unique Identifier
DTM: Dynamic Topic Model
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
MeSH: Medical Subject Heading
PHI: Public Health Index
ROI: Research Opportunity Index
SGML: Standard Generalized Markup Language
UMLS: Unified Medical Language System
USPC: United States Patent Classification

Edited by G Eysenbach; submitted 05.01.19; peer-reviewed by C Lin, M Spiliopoulou, H Tsai, S Berry, Y Ge; comments to author 30.01.19; revised version received 12.04.19; accepted 13.04.19; published 30.04.19.

Please cite as:
Huang M, Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Yao L
Technological Innovations in Disease Management: Text Mining US Patent Data From 1995 to 2017
J Med Internet Res 2019;21(4):e13316
URL: http://www.jmir.org/2019/4/e13316/
doi:10.2196/13316
PMID:31038462

Original Paper

Use of Electronic Health and Its Impact on Doctor-Visiting Decisions Among People With Diabetes: Cross-Sectional Study

Anne Helen Hansen[1,2], BSW, MD, PhD; Tor Claudi[3], MD; Eirik Årsand[4,5], Msc, PhD

[1]Centre for Quality Improvement and Development, University Hospital of North Norway, Tromsø, Norway
[2]Department of Community Medicine, Faculty of Health Sciences, UiT–The Arctic University of Norway, Tromsø, Norway
[3]Department of Medicine, Nordland Hospital, Bodø, Norway
[4]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
[5]Department of Clinical Medicine, UiT–The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Anne Helen Hansen, BSW, MD, PhD
Centre for Quality Improvement and Development
University Hospital of North Norway
PO Boks 35
Tromsø,
Norway
Phone: 47 91 61 96 55
Email: anne.helen.hansen@unn.no

Abstract

Background: Despite the increasing prevalence of diabetes and increasing use of electronic health (eHealth) among people with diabetes, little is known about the association between the use of eHealth and the use of provider-based health services.

Objective: The objective of this study was to investigate whether the use of eHealth might change patients' decisions regarding doctor-seeking behavior and whether information acquired from the internet was discussed with a doctor.

Methods: We used email survey data collected in 2018 from members of the Norwegian Diabetes Association (aged 18 to 89 years) diagnosed with diabetes. Using logistic regressions, we studied patients' internet-triggered changes in decisions regarding doctor visits; whether they discussed information from the internet with a doctor; and whether these topics were associated with gender, age, education, self-rated health, and self-reported anxiety/depression.

Results: Among the 895 informants, 75.4% (645/856) had never made an internet-triggered change of decision in any direction regarding visiting a doctor, whereas 16.4% (41/859) had decided to visit and 17.3% (148/856) had decided not to visit. The probability of changing decisions decreased with higher age and increased with the severity of self-reported anxiety/depression. Around half of the study participants (448/858, 52.2%) had never discussed information from the internet with a doctor. The probability of discussing internet information with a doctor was higher for those in bad/very bad self-rated health (odds ratio 2.12, CI 1.15-3.90) and for those with moderate self-reported anxiety/depression (odds ratio 2.30, CI 1.30-4.10).

Conclusions: Our findings suggest that using eHealth has a significant impact on doctor-visiting decisions among people with diabetes, especially among people aged 18 to 39 years and among those reporting anxiety/depression. It is of great importance that the information posted is of high quality and that the large differences between internet-users regarding age as well as mental and somatic health status are taken into account. More research is needed to confirm and further explore the findings of this study.

(J Med Internet Res 2019;21(4):e13678) doi:10.2196/13678

KEYWORDS

eHealth; internet; internet information; doctor-seeking behavior; cross-sectional study; diabetes; Norway

Introduction

Increasing Use of Electronic Health

The use of eHealth has developed and increased rapidly over the past decades. The term eHealth refers to “the transfer of health resources and health care by electronic means,” and internet use for health information is a part of this [1]. In this paper, we consider eHealth in the form of apps, search engines, social media, and video services. Of particular interest are patients with chronic disease, such as diabetes, who are in particular need of health information, monitoring, and follow-up of health and disease parameters.

Around 80% of the general population in the United States and Europe conduct health-related searches [2-5]. It was recently reported that 87% of Norwegians with type 1 diabetes (T1D) used eHealth in one or more forms, and 84% had used search engines (such as Google) sometimes or often during the previous year [6]. Most Norwegian households (98%) have internet access [7], 96% of the population aged 16 to 79 years have used the internet during the previous 3 months, and 90% use the internet every day [8].

Increasing Prevalence of Diabetes

The prevalence of diabetes is increasing worldwide and expected to rise to 642 million cases in 2040 [9]. Global prevalence in adults is estimated at 8.8% [9]. Around 245,000 persons have been diagnosed with diabetes in Norway, of whom around 28,000 have T1D [10]. Despite a decreasing trend in the incidence of type 2 diabetes (T2D) in Norway, the prevalence increased from 4.9% to 6.1% from 2009 to 2014 [11]. Most patients do not reach the combined national treatment targets for prevention of complications [12-14].

Relationship Between the Use of eHealth and Doctor-Visiting Behavior

A positive association between the use of eHealth and general practitioner (GP) visits has been reported [15]. Others have found an inverse association with the frequency of doctor visits [16,17]. eHealth might be used to decide about the need for seeing a doctor, seek information before the visit, or seek reassurance or additional information after the visit [6,18].

Some have suggested that the use of eHealth may postpone or replace medical consultations [19]. A French study found that around 30% of young eHealth users (aged 15 to 30 years) often used the internet instead of visiting a doctor [20]. In the same study, 88.6% reported that eHealth use did not change their consultation frequency, whereas 4.9% reported seeing a doctor more often and 6.5% less often. A German study found that frequent users of health services were 73% more likely to seek health information on the internet compared to nonusers [21]. This is in line with the illness behavior model [22], where people in poor health are more likely to seek disease-related information online and use health services more frequently. In a previous study among people with T1D, we found that the use of search engines was associated with more specialist visits, whereas GP visits were not associated with the use of eHealth [6]. Whether use of the internet for health information in a Norwegian population with diabetes might lead to or prevent doctor visits is unknown.

The use of eHealth often takes place without doctor involvement. Research from the United States reports that only 31% of mobile health (mHealth, which refers to mobile and wireless communication technologies to aid in health and health care) users prioritized their physician’s involvement [23]. To what extent diabetes patients in Norway discuss internet information with their doctor is unknown.

Due to the increased prevalence of diabetes and increased use of eHealth, doctor-visiting behavior related to the use of eHealth among people with chronic diseases such as diabetes is highly relevant for research. The use of eHealth among people with diabetes might have an impact on health information level, handling and care of health and disease, and doctor-seeking behavior and thus affects the overall important outcomes morbidity and mortality.

Aim

The aim of this study was to investigate whether the use of eHealth might lead to or prevent doctor visits and whether such changes in doctor-seeking behavior might be associated with gender, age, education, self-rated health, and self-reported anxiety/depression among people with diabetes. Furthermore, we aimed to investigate whether information from the internet was discussed in the clinical encounter and whether this was associated with gender, age, education, self-rated health, and self-reported anxiety/depression.

Methods

Data

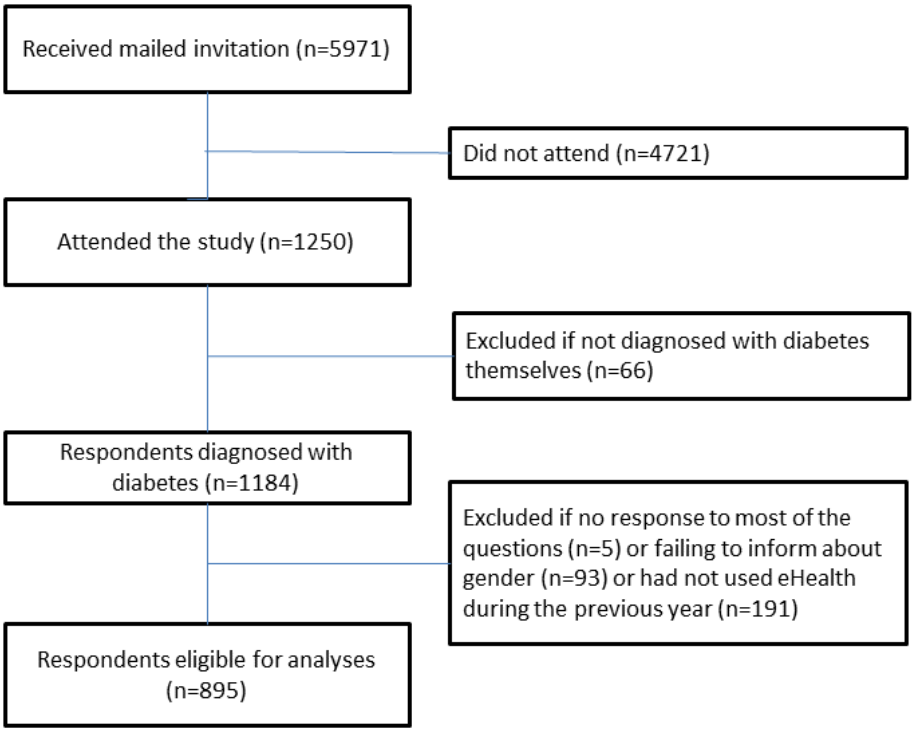
This cross-sectional study is part of the DIAcare project [24], which uses data obtained in 2018 from members of the Norwegian Diabetes Association (NDA). As of December 31, 2017, the organization had 33,908 members, of whom about 30% have T1D [25]. The Norwegian Centre for Research Data (NSD) distributed the invitations to a randomly selected sample of 5971 individuals with email addresses registered with NDA, who answered through NSD’s secure Web survey system. As described in our protocol paper, we planned to use data from the seventh Tromsø Study, conducted in 2015-2016 [26]. However, the Tromsø Study could not give us access, and we decided to develop a tailored questionnaire based on the specific objectives of our study using relevant questions from other published surveys [26,27].

Information about the study was posted together with the invitation. The questionnaire (Multimedia Appendix 1) included questions about health status, disease duration, severity and treatment of diabetes, use of and experiences with eHealth and health care services, and demographic and socioeconomic information. We reviewed and tested the questionnaire several times before distribution to the informants. Nonrespondents were given one reminder, sent by email 15 days after the first request.

Participants

The respondents could not fill in the questionnaire more than once. Starting from 1250 participants, we excluded the 66 individuals who had not been diagnosed with diabetes themselves (family members, health personnel, and others). We also excluded those who left out most of the questions (n=5) and those who did not give information about gender (n=93). Finally, we excluded those who had not used eHealth in the form of apps, search engines, social media, or video services "sometimes or often" during the previous year (n=191). The analyzed sample consisted of 895 respondents with diabetes (Figure 1).

Figure 1. Flowchart of study population.



Variables

We used 3 dependent variables developed from the following questions: Based on the information you have found on the internet, have you (1) decided to visit a doctor, when you would otherwise not have visited one, (2) decided not to visit a doctor, when you would otherwise have visited one, or (3) discussed the information with a doctor?

We dichotomized these variables by merging the original 4 response options into "never" and "once, sometimes or often" for two reasons. First, we were more interested in investigating if the participants had ever changed their doctor-seeking behavior or discussed internet information with a doctor than finding out how often they had done so. Second, merging the answering options into 2 groups gave an easier and more reader-friendly interpretation of the regression results.

Age was grouped in 20-year intervals. The 4 education categories were labeled low (primary/part of secondary school), middle (high school), high (college/university less than 4 years), and highest (college/university 4 years or more). Response options for self-rated health were excellent, good, fair, bad, and very bad. Due to low numbers in the very bad category (4 respondents), we merged the bad and very bad categories. Response options for self-reported degree of anxiety/depression were none, slight, moderate, severe, and extreme. We merged the severe and extreme categories due to few (4) respondents in the extreme category.

Analyses

We analyzed data by means of descriptive statistics and logistic regressions and tested correlations with the Spearman correlation test.

We constructed 3 regression models, with the 3 different dependent variables: "Decided to visit a doctor," "Decided not to visit a doctor," and "Discussed internet information with a doctor." The independent variables were gender, age, education, self-rated health, and self-reported degree of anxiety/depression. All the independent variables were introduced collectively into the multivariable models.

We subsequently introduced the response time variable into the regression models, comparing the late respondents with the early respondents, assuming that those who did not respond at first were more similar to nonrespondents [28].

We used 95% confidence intervals and set *P*<.05 as the significance level throughout the study. All analyses were accomplished using Stata version 14.2 (StataCorp LLC).

Ethics

The Regional Committee for Medical and Health Research Ethics found that an application for this project was not required according to the Norwegian Health Research Act (ref 2015/1779/REK nord). The data protection officer at the University Hospital of North-Norway approved the study (ref 2017/6579). The data bureau NSD received no information about the participants other than the email addresses.

Results

Participation

A total of 1250 persons aged 18 to 89 years answered the questionnaire, constituting a minimum response rate of 20.93%. We assume the real response rate to be higher since we had more than 400 bounce backs from email servers unable to deliver the invitation, and we do not know how many actually received the survey email. Eligible for analysis were the 895 persons who reported having been diagnosed with diabetes themselves.

Characteristics of the Participants

Mean age of participants was 53.5 years, 50.4 years for women and 56.3 years for men. Median age was 56 years. Mean disease duration was 13.6 (median 10) years.

The largest groups were made up of men (460/895, 51.4%), persons aged 60 years and over (369/895, 41.2%), persons with high education (265/832, 31.8%), diabetes duration of less than 10 years (295/892, 33.1%), good self-rated health (446/887, 50.3%), and no anxiety/depression (603/882, 68.4%). Most participants (718/859, 83.6%) had never made an internet-triggered decision to visit a doctor when they would otherwise not have visited one, whereas 16.4% (141/859) had made such a decision (Table 1). On the other hand, 82.7% (708/856) had never decided not to visit a doctor when they would otherwise have visited one, meaning that 17.3% (148/859) had made such a decision. When adding these 2 variables, we found that 75.4% (645/856) had never changed their decision regarding visiting a doctor in any direction based on information from the internet. These figures indicate that some had changed decisions in both directions.

A little more than half of the participants had never discussed information from the internet with a doctor (448/858, 52.2%).

Probability of an Internet-Triggered Change in the Decision to Visit or Not to Visit a Doctor

The probability of an internet-triggered decision to visit a doctor when one would otherwise not have visited one decreased with higher age and increased with the severity of self-reported anxiety/depression. People aged 60 years and over were significantly less likely to change their decision in this direction, compared with people aged 18 to 39 years (odds ratio [OR] 0.39, CI 0.23-0.67). The probability of changing the decision in this direction was more than 3 times higher for those with severe anxiety/depression (OR 3.2, CI 1.28-8.04), compared with no anxiety/depression. Gender, education, and self-rated health were not associated with deciding to visit a doctor based on information from the internet (Table 2).

The probability of an internet-triggered decision in the opposite direction (not to visit when one would otherwise have visited a doctor) followed the same pattern, with the exception that men were less likely to change their decision in this direction, compared with women (OR 0.63, CI 0.43-0.93). People aged 60 years and over were less likely to decide not to visit a doctor when they would otherwise have visited one, compared with people aged 18 to 39 years (OR 0.36, CI 0.22-0.59). Likewise, the probability of changing their decision in this direction was around 3 times higher for people with severe anxiety/depression (OR 2.97, CI 1.18-7.47), compared with no anxiety/depression. Education and self-rated health were not associated with deciding not to visit a doctor based on information from the internet.

Probability of Discussing Information From the Internet With a Doctor

People with bad/very bad self-rated health had a significantly higher probability of discussing information from the internet with a doctor, compared with those in excellent self-rated health (OR 2.12, CI 1.15-3.90). Discussing internet information with a doctor was associated with a moderate degree of self-reported anxiety/depression, compared with no anxiety/depression (OR 2.30, CI 1.30-4.10). Gender, education, and self-rated health were not associated with discussing information from the internet with a doctor (Table 3).

All findings presented in Tables 2 and 3 persisted after introduction of the response time variable into the regression models. There were no strong correlations (defined as Spearman rho>.5) between the independent variables. The strongest correlations were found for the variables self-rated health and self-rated anxiety/depression (rho=.3442). There were no other correlations above rho=.1471 (age and gender).

Table 1. Sample characteristics.

Characteristic	Value, n (%)
Gender (n=895)	
Female	435 (48.6)
Male	460 (51.4)
Age in years (n=895)	
18-39	180 (20.1)
40-59	346 (38.7)
60 and over	369 (41.2)
Education (n=832)[a]	
Low	74 (8.9)
Middle	241 (29.0)
High	265 (31.8)
Highest	252 (30.3)
Duration of diabetes, years (n=892)	
<10	295 (33.1)
10-19	253 (28.3)
20-29	170 (19.1)
30 and over	174 (19.5)
Self-rated health (n=887)	
Excellent	125 (14.1)
Good	446 (50.3)
Fair	224 (25.2)
Bad/very bad	92 (10.4)
Self-reported degree of anxiety/depression (n=882)	
None	603 (68.5)
Slight	181 (20.5)
Moderate	72 (8.2)
Severe	26 (2.9)
Based on information from the internet, have you decided to visit a doctor? (n=859)	
Never	718 (83.6)
Once	81 (9.4)
Sometimes	58 (6.8)
Often	2 (0.2)
Based on information from the internet, have you decided not to visit a doctor? (n=856)	
Never	708 (82.7)
Once	47 (5.5)
Sometimes	88 (10.3)
Often	13 (1.5)
Have you discussed information from the internet with a doctor? (n=858)	
Never	448 (52.2)
Once	132 (15.4)
Sometimes	254 (29.6)
Often	24 (2.8)

[a] Low (primary/part of secondary school), middle (high school), high (college/university less than 4 years), highest (college/university 4 years or more).

Table 2. Probability of changing the decision to visit or not to visit a doctor based on information from the internet.

Characteristics	Decided to visit a doctor when you would otherwise not have visited one (n=821)			Decided not to visit a doctor when you would otherwise have visited one (n=819)		
	OR[a]	*P* value	CI	OR	*P* value	CI
Gender						
Female	1.00	—[b]	—	1.00	—	—
Male	0.75	.15	0.50-1.11	0.63	.02	0.43-0.93
Age in years						
18-39	1.00	—	—	1.00	—	—
40-59	0.56	.02	0.34-0.92	0.43	.001	0.27-0.70
60 and over	0.39	.001	0.23-0.67	0.36	<.001	0.22-0.59
Education[c]						
Low	1.00	—	—	1.00	—	—
Middle	1.11	.79	0.50-2.45	1.05	.90	0.50-2.21
High	1.94	.09	0.90-4.17	1.10	.79	0.53-2.31
Highest	1.04	.92	0.47-2.31	1.28	.51	0.61-2.68
Self-rated health						
Excellent	1.00	—	—	1.00	—	—
Good	1.31	.44	0.66-2.58	1.35	.37	0.70-2.60
Fair	1.73	.14	0.83-3.59	1.58	.21	0.78-3.22
Bad/very bad	2.04	.09	0.89-4.66	1.85	.14	0.82-4.17
Self-reported degree of anxiety/depression						
None	1.00	—	—	1.00	—	—
Slight	1.85	.01	1.15-2.96	2.03	.002	1.29-3.21
Moderate	2.54	.005	1.33-4.83	2.27	.01	1.19-4.31
Severe	3.2	.01	1.28-8.04	2.97	.02	1.18-7.47

[a] OR: odds ratio.

[b] Not applicable (reference group).

[c] Low (primary/part of secondary school), middle (high school), high (college/university less than 4 years), highest (college/university 4 years or more).

Table 3. Probability of discussing information from the internet with a doctor (n=821).

Characteristic	OR[a]	*P* value	95% CI
Gender			
Female	1.00	—[c]	—
Male	1.01	.97	0.76-1.34
Age			
18-39	1.00	—	—
40-59	1.01	.95	0.68-1.51
60 and over	0.89	.56	0.51-1.32
Education[b]			
Low	1.00	—	—
Middle	0.8	.41	0.46-1.37
High	1.39	.23	0.81-2.37
Highest	1.13	.65	0.66-1.94
Self-rated health			
Excellent	1.00	—	—
Good	1.21	.38	0.79-1.85
Fair	1.27	.33	0.79-2.05
Bad/very bad	2.12	.02	1.15-3.90
Self-reported degree of anxiety/depression			
None	1.00	—	—
Slight	1.29	.17	0.90-1.85
Moderate	2.3	.005	1.30-4.10
Severe	1.66	.25	0.71-3.90

[a] OR: odds ratio.

[b] Low (primary/part of secondary school), middle (high school), high (college/university less than 4 years), highest (college/university 4 years or more).

[c] Not applicable (reference group).

Discussion

Principal Findings

Most study participants (645/856, 75.4%) had never changed their decision regarding visiting a doctor based on information from the internet, whereas some had decided to visit (141/859, 16.4%,) and/or not to visit (148/856, 17.3%). The probability of changing the decision decreased with higher age and increased with the severity of self-reported anxiety/depression. Around half of the study participants (448/858, 52.2%) had never discussed information from the internet with a doctor. The probability of discussing internet information with a doctor increased for those in bad/very bad self-rated health (compared with excellent health) and for those with moderate self-reported anxiety/depression (compared with no anxiety/depression).

Internet-Triggered Changes in Doctor-Visiting Decisions

Approximately 3 out of 4 participants in this study reported that they had never changed their decision to visit or not to visit a doctor based on information from the internet, whereas 16.4% had decided to visit and 17.3% had decided not to visit. It is worth noting that some reported changes in both directions, and that the 2 groups were about the same size (Table 1). A study among elderly internet users in the Netherlands (data collected in 2011) found that 48% had decided to go to the doctor and 24% had decided not to go because of internet information [29]. The percentage of participants who decided to visit a doctor was nearly 3 times higher in the Dutch study compared with our results. On the other hand, a French study among young internet users found that 4.9% visited a doctor more often and 6.5% less often based on information from the internet [20]. Around 90% of participants in studies from the United States, Japan, and France reported that use of the internet for health purposes did not change their health care–seeking behavior (United States 94%, Japan 88.9%, France 88.6%), meaning that 10% actually did [20,30,31].

Direct comparison between these studies is challenging due to methodological and cultural heterogeneity. Our study sampled people with diabetes, whereas the other studies were conducted in general populations. Data were collected in 2001-2002 (United States), 2007 (Japan), 2010 (France), and 2011

(Netherlands), whereas we collected our data in 2018. Our participants had a higher mean age compared with the studies from United States (42.5 years), Japan (46.4 years), and France (22.6 years) but a lower mean age than the Dutch study (72.0 years).

Nevertheless, our rate might suggest an increase in the impact of internet information on doctor-visiting decisions from 2001 to 2018. The internet has developed substantially and an increasing amount of health information has become available. Furthermore, use of this information has greatly increased, with elderly people as a fast-growing user group [2-5,32,33]. In this perspective, our finding that 25% did change doctor-visiting decisions based on internet information whereas only 10% did so in 2001 supports this trend. It is not possible to judge from this study whether people with diabetes are more or less likely to change their doctor-visiting behavior based on internet information compared with a general population. There is evidence that frequent users of health services, such as people with diabetes or other chronic diseases, are more likely to use the internet for health information compared with nonusers [6,21], which might explain some differences in the rates. On the other hand, a recent study found that demographic differences were more important than the presence or absence of chronic disease in this regard [34]. Other possible explanations might be cultural differences and differences in the level of exposure to internet information, the need for regular as well as irregular doctor visits, and the ability to apply internet information to one's own health situation.

Internet-Triggered Changes in Doctor-Visiting Decisions Decreased With Higher Age

Internet-triggered changes in doctor-visiting decisions in either direction decreased significantly with higher age (Table 2). Around 80% of young French adults (mean age 22.6 years) trusted health information from the internet [20], whereas around 40% of older people in the Netherlands (mean age 72.0 years) trusted this information source [27]. Greater trust in information found on the internet among younger people might partly explain our finding, making younger people more able to change doctor-visiting decisions after assessing internet information. Since younger people are more adapted to the internet [35] and the internet might be more tailored for younger users, information might be more easily perceived and transformed according to the individuals' needs, and thus easier to trust. This point might be reinforced by the lower education in older age groups [36].

Internet-Triggered Changes in Doctor-Visiting Decisions Increased With the Severity of Anxiety/Depression

In this study, an internet-triggered change in either direction in the decision to visit a doctor increased with the severity of self-reported anxiety/depression, whereas there was no association with self-rated health (Table 2). Others have found a positive relationship between searching the internet for health information and psychological distress [20] and health anxiety [18,37]. As health anxiety levels increase, the relationship between online health information seeking and visiting a doctor based on information found online also increases [38,39]. The internet has the potential to reduce as well as exacerbate health anxiety, and individuals with moderate to high levels of health anxiety experience more anxiety during and after online symptom checking, whereas individuals with low illness anxiety experienced relief [38,40,41]. Naturally, this might lead to changes in doctor-visiting decisions in both directions. Seeking help from the internet for anxiety/depression might partly be explained by the confidentiality of the internet and reduction of stigma. For the same reasons, people with anxiety/depression might avoid doctor visits, in particular if help and support is available online. There are strong indications that anxiety/depression is undertreated [42]; thus, sufficient and adequate treatment should be a concern for health care services and policy makers. Developing a variety of treatment options for people with self-reported anxiety/depression, tailored for individual needs whether online or face to face with a provider, should be a priority.

Internet-Triggered Decisions Not to Visit a Doctor Were Less Likely Among Men

Men were less likely to change their decision in the direction of not visiting a doctor when they would otherwise have visited one, compared with women. Women still tend to take care of children and other family members' health more than men do [3,20,35]. Many studies report that women use eHealth more than men do, even if results are not consistent [4]. More searches and findings of adequate health information online among women might be a possible explanation of our result, as women to a larger extent might find what they need for themselves or others' health issues on the internet and thus decide not to visit a doctor.

Approximately Half of Participants Never Discussed Internet Information With a Doctor

Approximately half of the study participants (448/858, 52.2%) had never discussed information from the internet with a doctor. Other studies found that 54% in the Netherlands [29], 69% in the United States [23], and 83.5% in Japan [31] had never discussed information obtained from the internet with health professionals. Discussing internet information in the clinical encounter does not seem common, despite some variation. Reported patient-experienced barriers to discussion of internet information include resistance from the physician; disapproval by the physician; fear of embarrassment and of criticizing, offending, insulting, or confronting the physician; and lack of time during the visits [43]. On the other hand, discussions are facilitated by encouragement from the doctor or from online advertisements, by the presence of family members in the consultation, and by higher self-rated ability to appraise internet information and one's own health status [43]. Some found that men, older people, people with more children under 18 years, and people in poorer health had higher probability of discussing internet information with the doctor [43]. As the internet has developed to be the new "first line" of health services, discussing information from the internet with a doctor (the next level) might not be necessary as long as internet information is considered understandable and sufficient and contributes in solving the patients' problem without a doctor's visit.

Higher Probability of Discussing Internet Information With a Doctor for People With Bad/Very Bad Health and Moderate Anxiety/Depression

The probability of discussing internet information with a doctor increased for those in bad/very bad self-rated health (compared with excellent health) and for those with moderate self-reported anxiety/depression (compared with no anxiety/depression). Discussing with a doctor presupposes doctor visits. People with bad/very bad self-rated health and moderate anxiety/depression are likely to meet this requirement, as doctor visit rates might be higher in these groups. We did not find an association between discussing internet information with a doctor for those with severe anxiety/depression, possibly due to a heavy disease burden that might hinder doctor visits [42,44] and probably decrease internet searching as well.

Implications and Further Research

Understanding patterns of health care–seeking behavior in the digital era is important for planning of health services for the population. Our study indicates that doctors should be aware of the need to discuss information from patients' internet use in the clinical encounter. Also, more effort should be put in providing high-quality updated online information for patients. Such information might give good advice regarding the need for doctor visits and thus lead to more tailored use of health care services. This might be particularly important for patients with anxiety/depression.

This study did not address any possible associations between changing doctor-seeking behavior and discussing internet findings with a doctor, which would be interesting for future research. Did those who decided to visit a doctor based on information from the internet discuss the information that changed their decision? Future research should also include investigations of specific webpages used for health information, patient rating of the information quality, associations with the use of health care services, and other impacts of internet use like, for instance, lifestyle changes.

Limitations and Strengths

This study had some shortcomings, which were explored in detail in our first publication in the DIAcare project [6]. Summing up, we discussed the low estimated participation rate (the main limitation); distribution of the questionnaire by email; recall bias; the validity of self-reported data; the cross-sectional study design; and participation related to gender, age, health, socioeconomic group, and interest in the subject studied. Our conclusion was that younger individuals might be overrepresented but that it is not possible to judge the magnitude or direction of a possible nonresponse bias because different factors might pull the tendency in different directions or level each other out. The low response rate is in itself not an indication of low representativeness, as nonresponse bias may be a problem even if response rates are high. We suggested that nonresponse bias posed a limited threat to our study's validity; however, generalization should be made with caution.

Worries, anxiety/depression, and emotional distress are not defined according to diagnostic manuals in this study and rely solely on self-report. Since many people with different kinds of psychological distress do not seek help and thus are undiagnosed [42], we consider self-report to be interesting as such in this field and do not think that this has distorted the validity of our results.

It should also be mentioned that this study does not address whether participants used the internet frequently or infrequently, which could have contributed to explaining the results. Likewise, we cannot exclude other unmeasured confounders of the reported associations, such as accessibility of doctor visits and doctor characteristics.

This study also has some strengths, which are similar to the strengths discussed in the first paper in this project [6]. The most important strength is the focus on a scarcely investigated research field. Other strengths are the detailed questionnaire specifically tailored to people with diabetes, the recruitment of participants from all of Norway, the inclusion of a wide age span of participants, and that we were able to analyze the data shortly after they were collected. Finally, the cooperation with NDA made real user participation possible in the design, execution, and implementation of the study.

Conclusions

Our findings suggest that the use of eHealth seems to have a significant impact on doctor- visiting decisions among people with diabetes, especially among younger people (aged 18 to 39 years) and those reporting anxiety/depression. It is of great importance that the information posted is of high quality and that the large differences between eHealth users regarding age as well as mental and somatic health status are taken into account. Furthermore, we conclude that around half of the participants did not discuss information from the internet with a doctor. More research is needed to confirm and further explore the findings of this study.

Acknowledgments

We thank the members of the NDA for their cooperation in the execution and performance of this study. Without their generous contribution, this study could not have been realized. We also thank the Northern Norway Health Authorities and the University Hospital of North Norway, who provided funding for this research and approved the changes regarding data collection that had to be made in order to complete the study.

Authors' Contributions

All authors contributed to the design and execution of the study. AHH drafted the protocol and the manuscript. All authors contributed with improvements and critical revisions and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire.

[DOCX File, 14KB - jmir_v21i4e13678_app1.docx]

References

1. World Health Organization: eHealth. URL: http://www.who.int/ehealth/en/ [accessed 2018-06-16] [WebCite Cache ID 70DfcHSuE]
2. Wangberg SC, Andreassen HK, Prokosch H, Santana SMV, Sørensen T, Chronaki CE. Relations between Internet use, socio-economic status (SES), social support and subjective health. Health Promot Int 2008 Mar;23(1):70-77 [FREE Full text] [doi: 10.1093/heapro/dam039] [Medline: 18083686]
3. Fox S, Duggan M. Health Online 2013. Washington: Pew Internet and American Life Project URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2019-04-17] [WebCite Cache ID 77hjyUxQ9]
4. Hong YA, Cho J. Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]
5. Wangberg SC, Sorensen T, Andreassen HK. [Prosjektrapport e-helse i Norge 2013]. URL: https://ehealthresearch.no/files/documents/Prosjektrapporter/NST-rapport_2014-02_e-helse_i_Norge_2013.pdf [accessed 2018-06-16] [WebCite Cache ID 70DhrW5OT]
6. Hansen AH, Broz J, Claudi T, Årsand E. Relations between the use of electronic health and the use of general practitioner and somatic specialist visits in patients with type 1 diabetes: cross-sectional study. J Med Internet Res 2018 Nov 07;20(11):e11322 [FREE Full text] [doi: 10.2196/11322] [Medline: 30404766]
7. [Norsk Mediebarometer]. URL: http://medienorge.uib.no/statistikk/medium/ikt/347 [accessed 2019-01-31] [WebCite Cache ID 75pwPjShM]
8. [Use of ICT in households]. URL: https://www.ssb.no/teknologi-og-innovasjon/artikler-og-publikasjoner/ni-av-ti-surfer-pa-nettet-hver-dag [accessed 2019-01-31] [WebCite Cache ID 75pxg0TzO]
9. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
10. [Public Health in Norway 2018].: Norwegian Institute of Public Health URL: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2018/helsetilstanden-i-norge-2018.pdf [accessed 2018-06-16] [WebCite Cache ID 70DeT070T]
11. Ruiz PLD, Stene LC, Bakken IJ, Håberg SE, Birkeland KI, Gulseth HL. Decreasing incidence of pharmacologically and non-pharmacologically treated type 2 diabetes in Norway: a nationwide study. Diabetologia 2018 Dec;61(11):2310-2318 [FREE Full text] [doi: 10.1007/s00125-018-4681-4] [Medline: 29995214]
12. Jenssen TG, Tonstad S, Claudi T, Midthjell K, Cooper J. The gap between guidelines and practice in the treatment of type 2 diabetes: a nationwide survey in Norway. Diabetes Res Clin Pract 2008 May;80(2):314-320. [doi: 10.1016/j.diabres.2007.12.025] [Medline: 18279994]
13. Mouland G. [Diabetes in general practice—were treatment goals reached?]. Tidsskr Nor Laegeforen 2014 Jan 28;134(2):168-172 [FREE Full text] [doi: 10.4045/tidsskr.13.0375] [Medline: 24477150]
14. Cooper JG, Claudi T, Thordarson HB, Løvaas KF, Carlsen S, Sandberg S, et al. Treatment of type 1 diabetes in the specialist health service—data from the Norwegian Diabetes Register for Adults. Tidsskr Nor Laegeforen 2013 Nov 12;133(21):2257-2262 [FREE Full text] [doi: 10.4045/tidsskr.13.0153] [Medline: 24226332]
15. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
16. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15 Suppl 3:279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]
17. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]
18. 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]

19. Fox S. The social life of health information, 2011. Washington: Pew Internet and American Life Project URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2019-04-17] [WebCite Cache ID 77hkth3PU]
20. 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;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
21. 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;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
22. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10:1. [doi: 10.1177/1460458215595851] [Medline: 26261218]
23. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin L, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788 [FREE Full text] [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
24. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen O, Wangberg SC, et al. The use of eHealth and provider-based health services by patients with diabetes mellitus: protocol for a cross-sectional study. JMIR Res Protoc 2016 Oct 31;5(4):e207 [FREE Full text] [doi: 10.2196/resprot.6529] [Medline: 27799136]
25. [Årsberetning 2017].: Norwegian Diabetes Association URL: https://issuu.com/knutjarle/docs/a_rsberetning_2017_ [accessed 2018-06-16] [WebCite Cache ID 70DkKIFTA]
26. The Tromsø Study Website. URL: http://tromsoundersokelsen.uit.no/tromso/ [accessed 2018-06-16] [WebCite Cache ID 70Dkah0hU]
27. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
28. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA 2012 May 2;307(17):1805-1806. [doi: 10.1001/jama.2012.3532] [Medline: 22550194]
29. Medlock S, Eslami S, Askari M, Sent D, de Rooij SE, Abu-Hanna A. The consequences of seniors seeking health information using the internet and other sources. Stud Health Technol Inform 2013;192:457-460. [Medline: 23920596]
30. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
31. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
32. Bujnowska-Fedak MM. Trends in the use of the Internet for health purposes in Poland. BMC Public Health 2015;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
33. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
34. Asan O, Cooper II F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for health information technologies among US adults: analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
35. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
36. Level of education in the population.: Statistics Norway URL: https://www.ssb.no/utniv/ [accessed 2019-02-01] [WebCite Cache ID 75rcSjTbT]
37. 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 Dec 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
38. 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]
39. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
40. Singh K, Fox JRE, Brown RJ. Health anxiety and Internet use: a thematic analysis. J Psychosoc Res Cyberspace 2016;10(2):1. [doi: 10.5817/CP2016-2-4]
41. 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]
42. Hansen AH, Høye A. Gender differences in the use of psychiatric outpatient specialist services in Tromsø, Norway are dependent on age: a population-based cross-sectional survey. BMC Health Serv Res 2015 Oct 22;15:477 [FREE Full text] [doi: 10.1186/s12913-015-1146-z] [Medline: 26493268]

43. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
44. Hansen AH, Halvorsen PA, Ringberg U, Førde OH. Socio-economic inequalities in health care utilisation in Norway: a population based cross-sectional survey. BMC Health Serv Res 2012 Sep 25;12:336 [FREE Full text] [doi: 10.1186/1472-6963-12-336] [Medline: 23006844]

Abbreviations

GP: general practititioner
NDA: Norwegian Diabetes Association
NSD: Norwegian Centre for Research Data
OR: odds ratio
T1D: type 1 diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 11.02.19; peer-reviewed by S Oser, H Fu, A Fioravanti, C Liang; comments to author 02.03.19; revised version received 08.04.19; accepted 08.04.19; published 26.04.19.

Please cite as:
Hansen AH, Claudi T, Årsand E
Use of Electronic Health and Its Impact on Doctor-Visiting Decisions Among People With Diabetes: Cross-Sectional Study
J Med Internet Res 2019;21(4):e13678
URL: http://www.jmir.org/2019/4/e13678/
doi:10.2196/13678
PMID:31025957

Original Paper

Face-to-Face Versus Video Assessment of Facial Paralysis: Implications for Telemedicine

Jian Rong Tan[1*], BAppSc; Susan Coulson[1*], GradCertHigherEd, BAppSc, MAppSc, PhD; Melanie Keep[2*], BPsych, MEd, PhD

[1]Discipline of Physiotherapy, Faculty of Health Sciences, University of Sydney, Lidcombe, Australia
[2]Faculty of Health Sciences, University of Sydney, Lidcombe, Australia
[*]all authors contributed equally

Corresponding Author:
Susan Coulson, GradCertHigherEd, BAppSc, MAppSc, PhD
Discipline of Physiotherapy
Faculty of Health Sciences
University of Sydney
Cumberland Campus, Building O
75 East Street
Lidcombe, 2141
Australia
Phone: 61 93519416
Email: susan.coulson@sydney.edu.au

Abstract

Background: Patients with facial nerve paralysis (FNP) experience challenges in accessing health care that could potentially be overcome by telemedicine. However, the reliability of telemedicine has yet to be established in this field.

Objective: This study aimed to investigate the consistency between face-to-face and video assessments of patients with FNP by experienced clinicians.

Methods: A repeated-measures design was used. A total of 7 clinicians assessed the FNP of 28 patients in a face-to-face clinic using standardized grading systems (the House-Brackmann, Sydney, and Sunnybrook facial grading systems). After 3 months, the same grading systems were used to assess facial palsy in video recordings of the same patients.

Results: The House-Brackmann system in video assessment had excellent reliability and agreement (intraclass correlation coefficient [ICC]=0.780; principal component analysis [PCA]=87.5%), similar to face-to-face assessment (ICC=0.686; PCA=79.2%). Reliability of the Sydney system was good to excellent, with excellent agreement face-to-face (ICC=0.633 to 0.834; PCA=81.0%-95.2%). However, video assessment of the cervical branch and synkinesis had fair reliability and good agreement (ICC=0.437 to 0.597; PCA=71.4%), whereas that of other branches had good to excellent reliability and excellent agreement (ICC=0.625 to 0.862; PCA=85.7%-100.0%). Reliability of the Sunnybrook system was poor to fair for resting symmetry (ICC=0.195 to 0.498; PCA=91.3%-100.0%) and synkinesis (ICC=−0.037 to 0.637; PCA=69.6%-87.0%) but was good to excellent for voluntary movement (ICC=0.601 to 0.906; PCA=56.5%-91.3%) in face-to-face and video assessments. Bland-Altman plots indicated normal limits of agreement within ±1 between face-to-face and video-assessed scores only for the temporal and buccal branches of the Sydney system and for resting symmetry in the Sunnybrook system.

Conclusions: Video assessment of FNP with the House-Brackmann and Sunnybrook systems was as reliable as face-to-face but with insufficient agreement, especially in the assessment of synkinesis. However, video assessment does not account for the impact of real-time interactions that occur during tele-assessment sessions.

(J Med Internet Res 2019;21(4):e11109) doi:10.2196/11109

KEYWORDS

telemedicine; eHealth; facial paralysis; facial nerve; outcome assessment (health care); video recording; facial expression; smiling; quality of life; Bell palsy

Introduction

Background

Facial nerve paralysis (FNP) occurs when there is damage to the seventh cranial nerve, the facial nerve. This affects an individual's ability to express emotions via facial movements (eg, frowning, smiling, or pouting), thereby compromising interpersonal communication [1-4]. The effects of FNP are not only limited to physical and aesthetic deficits caused by weakness, synkinesis (involuntary muscle movements accompanying voluntary movements), and asymmetry of the resting and moving face [1-3,5-10]. FNP is also associated with functional effects such as oral incontinence, speech deficits, dry eye, and subsequent corneal damage and psychosocial effects such as decreased self-esteem, psychological distress, depression, and reduced quality of life [1-3,5-10]. It is, therefore, desirable that individuals with FNP receive appropriate help for their condition to reduce deficits and improve functional outcomes.

Management of facial nerve disorders is a highly specialized field with a limited number and availability of practitioners. For example, in Australia, there is only 1 major facial nerve center—the Sydney Facial Nerve Clinic (SFNC). In addition to limited access to relevant health professionals, individuals with FNP face common challenges to health care access, which include the direct and indirect cost of care, distance, transportation, mobility, and time constraints [11-15].

One potential way of improving access to care for patients who experience FNP is telemedicine. Telemedicine is health care delivered via telecommunications technology over a distance [11,15-18], allowing improved access to care in various physical and mental health conditions by overcoming geographical and temporal barriers [11,13-21]. Telemedicine has also been associated with increased access and compliance to pulmonary rehabilitation [22] and reduced hospitalization rates, emergency department visits, and length of hospital stay for patients with chronic obstructive pulmonary disease [23]. Patients have reported similar or better quality of life compared with usual care [23]. In addition, internet-based interventions have been found to significantly decrease pain intensity, depression, anxiety, and stress and increase quality of life compared with in-person psychotherapy in patients with medically unexplained pain [20], in youth with depression and anxiety [21], in military personnel with post-traumatic stress disorder and depression [17,24], and in patients with neurofibromatosis [15].

Limitations of telemedicine include technical difficulties such as poor video resolution, connectivity, and sound issues [17,18,20,21,25]. These could potentially compromise the quality of the service provided and are commonly cited reasons for not implementing telehealth [17,18,20,21,25].

Although telemedicine has not yet been formally studied in FNP, it has, anecdotally, been implemented by some clinicians to provide services for patients experiencing geographical barriers to treatment. The above evidence suggests that telemedicine may play a role in increasing access to medical care for patients with FNP. As such, studies are required to establish the reliability of telemedicine as compared with face-to-face health care of FNP.

Tele-assessment of facial function using the appropriate facial grading systems is an integral part of the management of FNP as FNP management requires initial and ongoing assessment of patients' facial function. The clinician's overall impression of the patient is taken as standard and is assumed to represent the degree of abnormality of facial function [1,2,4-6,8,26-36]. This enables clinicians to keep track of patients' recovery and evaluate the efficacy of treatment [4,28,31,37]. FNP is assessed using grading scales that typically measure patients' facial function, grade symmetry of the face at rest, and measure displacement of facial features during voluntary movement as well as in the presence of synkinesis. Previous research has shown that greatest displacement during facial expressions occurs in the vertical axis of the frontal plane, followed by the anterior-posterior axis of the sagittal plane and the horizontal axis of the frontal plane [27]. This raises the question of whether tele-assessment of the patient, which would present the patient in a 2-dimensional (2D) view, would be as reliable and valid as a face-to-face assessment where patients are presented in a 3-dimensional (3D) view.

Objectives

Despite the significant body of literature on the applications of telemedicine, there has been a lack of research conducted on its use for the management of FNP. This exploratory study aims to compare the reliability of 3 commonly used FNP grading systems when administered via static video and face-to-face. Face-to-face assessment presents the patient in a 3D view, whereas the frontal video shows the patient in a 2D view that would not take into account the anterior-posterior axis of the sagittal plane. Therefore, it was hypothesized that reliability across the 3 grading systems would differ between the face-to-face and video assessments in movements where there was significant anterior-posterior axis displacement, for example, the pout, smile, and snarl.

Methods

Study Design

A repeated-measures design was undertaken in which the face-to-face and video assessments were conducted on the same patients and assessed by the same assessors. The video assessment was performed on a static prerecorded video of the patients taken before the face-to-face assessment. There was a 3-month delay between the face-to-face and video assessments to minimize retention of gradings used by the assessors. The 3-month period was based on a similar study by Banks et al [31], which compared in-person and video assessment of facial mimetic function using the newly developed eFACE facial grading system.

Participants

Participants comprised clients at the SFNC who presented with a diagnosed FNP between August 2016 and March 2017. Assessment was conducted in a standard, routine manner, and the video assessment segment of the study did not require additional involvement by the participants, nor did it preclude them from receiving treatment or other appropriate interventions. Ethical approval was granted by the Ethics Review Committee (RPAH Zone) of the Sydney Local Health District (Protocol No X17-0013 ERC).

Procedure

The study took place between August 2016 and March 2017. Patients of the SFNC are routinely video-recorded while performing a protocol of facial movements before receiving treatment. Videos were recorded in the frontal view and included movements such as eyebrow raise, eye closure, smile, snarl, and pucker. This is not a recording of a live videoconference. Static videos were used as it allowed researchers to control for individual differences in how clinicians guided patients through assessment protocols in a live videoconferencing session. In addition, it was important that the patients' state of facial paralysis at the face-to-face and video assessments were identical. Some recovery of facial function might have occurred if the videoconferencing assessment was conducted after the face-to-face assessment.

After the recording is taken, participants commence their in-person clinic session where the face-to-face assessment was conducted. Moreover, 3 months later, each assessor rated the video footage taken at the initial face-to-face session. The order in which the videos were shown was randomized to control for rater fatigue. Each assessor graded each patient face-to-face with the House-Brackmann system, the Sunnybrook system, and the Sydney system and then again in the video assessment.

Face-to-face assessment was the standard against which the video assessment was compared. This process enabled inferences to be made about tele-assessment as a potential substitute for face-to-face assessment.

House-Brackmann Facial Grading System

The House-Brackmann facial grading system is the most commonly used system to measure FNP [2-5,8,26-28,30,31,33,34,37]. It is a gross scale ranked from 1 to 6, with each grade giving an overall impression of facial function, resting symmetry, and synkinesis [4,5,28,34]. However, the House-Brackmann system is limited in that it does not allow for regional assessment, the range of scores does not reflect the clinically important change, and lacks sufficient classification of synkinesis [8,26,30,32,34-36]. This has led to the administration of a range of alternative systems being used by practitioners around the world.

Sydney Facial Grading System

The Sydney Facial Grading System, which assesses voluntary movement of the 5 branches of the facial nerve and overall synkinesis, has also been used regionally in Australia [1,6,26] and reported in the International Facial Nerve Symposium conference proceedings over a 20-year period.

Sunnybrook Facial Grading System

The Sunnybrook facial grading system is a regional weighted system that assesses resting symmetry, voluntary movements, and synkinesis of the face, after which a composite score on a 100-point scale is computed [28,34]. It grades patients in a more objective and continuous manner than the House-Brackmann system, and each component of the system is sensitive to change and contributes to a change in the composite score [1,3,8,26,28,30,32,34,35].

Analyses

Participants' data were deidentified before analyses. Quantitative analyses were conducted using the Statistical Package for the Social Sciences. Descriptive statistics were used to summarize demographic information. A significance level of $P<.05$ and corresponding 95% CIs were used for all inferential statistics.

Agreement between face-to-face and video House-Brackmann grades, Sydney scores, and Sunnybrook scores was assessed using the 1-way random, single measure intraclass correlation coefficient (ICC, 1,1), and 2-tailed repeat measures *t* tests. ICC (1,1) was selected because each participant was assessed by a different set of assessors. For each participant, scores given by 3 randomly selected assessors were used to compute the ICC; hence, the rater was a random effect [38,39]. Portney and Watkins [38] noted that "ICC ranges between 0.00 and 1.00, with values closer to 1.00 represent stronger reliability," but there were "no standard values for acceptable reliability using the ICC," and that the researcher should determine the level of reliability needed to justify the use of the tool being assessed. On the basis of the study by Banks et al [31], we defined reliability as poor for ICC values less than 0.40, fair for values between 0.40 and 0.59, good for values between 0.60 and 0.74, and excellent for values 0.75 and above. The 95% CI indicates that there is a 95% chance that any score would lie within the range [39].

As the ICC could be affected by outlying data, the percentage exact agreement (PEA) and percentage close agreement (PCA) were calculated to make up for the potential deficiency [40]. PEA is the proportion of cases where all 3 assessors gave the same grade for each participant. PCA is similar to PEA but includes cases where there was a maximum difference of 1 grade between assessors [40]. Agreement of PEA with ICC suggests the data are sound [40]. However, as a 1-grade and 17-point difference between assessors had been determined to be reasonable for the grading systems and the Sunnybrook composite score, respectively [4,32], we will be comparing the ICC with the PCA instead. Similar to the ICC, agreement was defined as poor for PCA values less than 40%, fair for values between 40% and 59%, good for values between 60% and 74%, and excellent for PCA values 75% and above.

Repeat measures *t* tests were administered on a per-protocol analysis. Missing data (ie, no matched pairs of face-to-face and video assessment grades) were discarded.

Bland-Altman plots were constructed where differences between face-to-face and video assessments were plotted against the means of the 2 assessment modes to determine agreement between the two and whether video assessment can substitute face-to-face assessment with an acceptable degree of error [38,41,42]. Horizontal lines were drawn at the mean difference and at the upper and lower limits of agreement, which are defined as the mean difference plus and minus 1.96 times the SD of the differences, respectively [38,41,42]. Assuming normal distribution, between the limits of agreement is an agreement interval within which 95% of the differences of video assessment fall, compared with face-to-face assessment [38,43]. The Bland-Altman method only defines the intervals of agreement but not whether those limits are acceptable [43]. We have, therefore, defined a difference of 1 grade to be acceptable between House-Brackmann grades and components of the Sydney and Sunnybrook systems and a difference of 17 points in the Sunnybrook composite score.

Results

Participants

A total of 28 participants and 7 assessors took part in the study. The mean (SD) age of the 28 patient participants (12 males and 16 females) was 41.7 (15.4) years, with a mean (SD) of 6.5 (6.9) years since FNP onset. There has been no prior research comparing face-to-face and video assessments of FNP, so the sample size (n=28) was based on a related study with 21 participants where there was sufficient power to detect differences between the House-Brackmann, Sydney, and Sunnybrook systems [26]. The etiologies of the FNP are presented in Table 1.

The assessors comprised ear, nose, and throat specialists; head and neck surgeons; and plastic surgeons who regularly attended the SFNC and had at least 10 years of experience in managing facial nerve disorders.

Face-to-Face Versus Video-Assessed House-Brackmann Grades

The House-Brackmann system administered face-to-face had good reliability and excellent agreement, whereas that administered over a video had excellent reliability and agreement (Table 2). Significant between-subject effects are denoted by significant *P* value in an *F* test with true value 0. When the between-subject effect is significant (*P*<.05), participants are then different from each other, which is a necessary condition for reliability testing [38].

The Bland-Altman plot indicated that the mean difference between House-Brackmann grades given in the face-to-face and video assessments was 0.05, but limits of agreement were between −1.79 to +1.89.

Face-to-Face Versus Video-Assessed Sydney System Scores

Reliability in assessing the 5 branches of the facial nerve face-to-face using the Sydney system was good to excellent, with excellent agreement (Table 3). However, video assessment of the cervical branch and synkinesis had only fair reliability and good agreement, whereas video assessment of the other branches had similarly good to excellent reliability and excellent agreement.

Inspection of the Bland-Altman plots showed that the mean differences between face-to-face and video assessments were close to 0 when assessing all facial nerve branches using the Sydney system. However, limits of agreement between the 2 modes of assessment were only close to ±1 in assessment of the temporal and buccal branches, and a patient could potentially be graded more than 1 grade apart between the face-to-face and video assessments of the zygomatic, marginal mandibular, and cervical branches and synkinesis.

Table 1. The etiology of facial nerve paralysis in study participants.

Diagnoses	Statistics, n (%)	Type
Trauma or iatrogenic	13 (46)	Removal of acoustic neuromas, parotid gland tumors, hemangioma, and basal cell carcinoma
Bell’s palsy	5 (18)	Idiopathic
Tumor	4 (14)	Facial nerve neuroma
Herpes zoster oticus	3 (11)	Viral (Ramsay Hunt Syndrome)
Congenital	2 (7)	Developmental
Trauma or accidental	1 (4)	Gunshot wound

Table 2. Reliability and agreement of the House-Brackmann scale across assessment modes.

Assessment mode	Intraclass correlation coefficient (95% CI)	*P* value	Percentage exact agreement (%)	Percentage close agreement (%)
Face-to-face	0.686 (0.488-0.835)	<.001	41.7	79.2
Video	0.780 (0.623- 0.889)	<.001	25.0	87.5

Table 3. Reliability and agreement of Sydney system facial nerve branches across assessment modes.

Sydney system facial nerve branches	Assessment mode	Intraclass correlation coefficient (95% CI)	*P* value	Percentage exact agreement (%)	Percentage close agreement (%)
Temporal	Face-to-face	0.789 (0.624-0.900)	<.001	52.4	95.2
	Video	0.857 (0.734-0.934)	<.001	66.7	90.5
Zygomatic	Face-to-face	0.633 (0.402-0.813)	<.001	33.3	90.5
	Video	0.625 (0.392-0.809)	<.001	42.9	85.7
Buccal	Face-to-face	0.834 (0.695-0.922)	<.001	61.9	95.2
	Video	0.862 (0.742-0.936)	<.001	61.9	100.0
Marginal mandibular	Face-to-face	0.794 (0.630-0.902)	<.001	52.4	85.7
	Video	0.758 (0.576-0.883)	<.001	33.3	85.7
Cervical	Face-to-face	0.790 (0.624-0.900)	<.001	42.9	81.0
	Video	0.597 (0.356-0.792)	<.001	38.1	71.4
Synkinesis	Face-to-face	0.709 (0.504-0.857)	<.001	57.1	90.5
	Video	0.437 (0.172-0.687)	<.001	42.9	71.4

Face-to-Face Versus Video-Assessed Sunnybrook Scores

The reliability of face-to-face assessment using the Sunnybrook system was similar to video assessment across most parameters. The reliability was generally poor to fair when assessing resting symmetry and synkinesis but was good to excellent when assessing voluntary movement (Table 4). Agreement was generally good to excellent across parameters for both face-to-face and video assessments. The Sunnybrook composite score had excellent reliability in both face-to-face and video assessments.

The *F* test for between-subjects effect was not significant in the face-to-face assessment of resting symmetry of the eye and in both modes of assessment of synkinesis in forehead wrinkle, indicating no significant difference amongst patients' resting symmetry of the eye and presence of synkinesis in the forehead wrinkle.

No PCAs were calculated for the weighted total resting symmetry and weighted total voluntary movement scores. As the raw scores were multiplied by 5 and 4, respectively, to determine the weighted total scores, there would not be cases where scores differed by only 1 point. No PEAs and PCAs were calculated as well for the composite score, which was derived from a formula in which resting symmetry, voluntary movement, and synkinesis were weighted differently. A total of 2 patients with similar Sunnybrook composite scores might, therefore, present differently, with 1 patient scoring better in resting symmetry and the other scoring better in voluntary movement or synkinesis, so agreement in scores might not necessarily reflect similar facial function anyway.

The Bland-Altman plots demonstrated mean differences of close to 0 between face-to-face and video assessments of resting symmetry when using the Sunnybrook system. Limits of agreement between the 2 modes of assessment of resting symmetry were overall close to ±1. Assessment of voluntary movement with the Sunnybrook system saw mean differences of close to 0 between face-to-face and video assessments. Limits of agreement between the 2 modes of assessment of voluntary movement were largely more than ±1 and closer to ±2. Assessment of synkinesis with the Sunnybrook system saw mean differences of close to 0 between face-to-face and video assessments. Limits of agreement between the 2 modes of assessments of synkinesis were overall more than ±1 and closer to ±2. The Bland-Altman plots also indicated a mean difference of 1.52 points in the Sunnybrook composite score between face-to-face and video assessments, which was within the previously defined reasonable limits of 17 points but with limits of agreement greater than ±17 points. Missing data were removed and not included in the results.

Table 4. Reliability and agreement of Sunnybrook system parameters across assessment modes. No percentage close agreement (PCA) was calculated for the weighted total resting symmetry and weighted total voluntary movement scores because raw scores were multiplied by 5 and 4 respectively to determine the weighted total scores; there would not be cases where scores differed by only 1 point. No PEA and PCAs were calculated for the composite score, which was derived from a formula in which resting symmetry, voluntary movement, and synkinesis were weighted differently.

Sunnybrook system components	Assessment mode	Intraclass correlation coefficient (95% CI)	*P* value	Percentage exact agreement (%)	Percentage close agreement (%)
Resting symmetry					
Eye	Face-to-face	0.195 (−0.045 to 0.480)	.06	43.5	100.0
	Video	0.367 (0.113 to 0.625)	<.001	73.9	100.0
Cheek	Face-to-face	0.332 (0.078 to 0.597)	.01	43.5	91.3
	Video	0.498 (0.250 to 0.719)	<.001	52.2	95.7
Mouth	Face-to-face	0.235 (−0.011 to 0.516)	.03	43.5	100.0
	Video	0.314 (0.061 to 0.582)	.01	47.8	100.0
Total x 5	Face-to-face	0.548 (0.308 to 0.753)	<.001	34.8	—[a]
	Video	0.544 (0.303 to 0.750)	<.001	26.1	—
Voluntary movement					
Forehead wrinkle	Face-to-face	0.859 (0.74 to 0.932)	<.001	43.5	82.6
	Video	0.906 (0.82 to 0.955)	<.001	60.9	87.0
Eye closure	Face-to-face	0.601 (0.367 to 0.790)	<.001	22.7	77.3
	Video	0.688 (0.481 to 0.842)	<.001	18.2	77.3
Smile	Face-to-face	0.741 (0.561 to 0.869)	<.001	26.1	78.3
	Video	0.818 (0.677 to 0.911)	<.001	34.8	91.3
Snarl	Face-to-face	0.707 (0.512 to 0.850)	<.001	21.7	69.6
	Video	0.806 (0.659 to 0.905)	<.001	39.1	82.6
Lip pucker	Face-to-face	0.690 (0.489 to 0.840)	<.001	34.8	56.5
	Video	0.826 (0.691 to 0.915)	<.001	69.6	82.6
Total x 4	Face-to-face	0.812 (0.668 to 0.908)	<.001	8.7	—
	Video	0.912 (0.834 to 0.958)	<.001	21.7	—
Synkinesis					
Forehead wrinkle	Face-to-face	−0.006 (−0.203 to 0.276)	.50	52.2	73.9
	Video	−.037 (−0.225 to 0.241)	.60	43.5	69.6
Eye closure	Face-to-face	0.248 (0.000 to 0.527)	.03	56.5	73.9
	Video	0.299 (0.047 to 0.570)	.01	34.8	73.9
Smile	Face-to-face	0.562 (0.325 to 0.762)	<.001	52.2	69.6
	Video	0.527 (0.284 to 0.739)	<.001	52.2	78.3
Snarl	Face-to-face	0.475 (0.225 to 0.703)	<.001	52.2	73.9
	Video	0.564 (0.327 to 0.763)	<.001	43.5	78.3
Lip pucker	Face-to-face	0.459 (0.208 to 0.692)	<.001	43.5	73.9
	Video	0.637 (0.418 to 0.809)	<.001	39.1	87.0
Total	Face-to-face	0.534 (0.292 to 0.744)	<.001	34.8	52.2
	Video	0.475 (0.225 to 0.703)	<.001	30.4	39.1
Composite	Face-to-face	0.811 (0.666 to 0.907)	<.001	—	—
	Video	0.845 (0.720 to 0.925)	<.001	—	—

[a] Missing data.

Discussion

Principal Findings

Reliability of the House-Brackmann, Sydney, and Sunnybrook systems was largely similar between face-to-face assessment and video assessment of FNP, despite the 3D nature of face-to-face assessment versus the 2D nature of video assessment. There was, however, poor reliability of assessment of synkinesis.

Face-to-Face Versus Video-Assessed House-Brackmann Grades

Reliability of the House-Brackmann scale was good in the face-to-face assessment and excellent in the video assessment, with excellent agreement within both modes of assessment. The high level of reliability supports the findings of Evans et al [4], House and Brackmann [33], Coulson et al [26], and Kanerva et al [32]. The lack of an anterior-posterior axis in the video assessment did not appear to affect reliability of the House-Brackmann scale negatively, as implied by the higher ICC value than in face-to-face assessment.

However, the Bland-Altman plot indicated that patients could potentially be given more than 1 grade difference between the face-to-face and video assessments. One grade of difference reflects significant variation in facial function; hence, using the House-Brackmann scale in tele-assessment may give the clinician an erroneous impression of the patient’s condition. Therefore, although the reliability of the House-Brackmann scale indicated that its implementation in a video assessment may yield comparable grades to face-to-face assessment, potential large differences between the 2 assessment modes meant that video assessment using the House-Brackmann scale cannot necessarily substitute face-to-face assessment.

Face-to-Face Versus Video-Assessed Sydney System Scores

Reliability in assessing the 5 branches of the facial nerve face-to-face using the Sydney system was good to excellent, with excellent agreement. Reliability and agreement were largely similar between face-to-face and video assessments, except for video assessment of the cervical branch and synkinesis, which were less reliable and had a lower agreement than the face-to-face assessment. This supports Coulson et al’s [26] findings of good reliability in assessing function of the facial nerve branches, although the ICC values reported were slightly lower, and poor reliability in assessment of synkinesis. The higher ICC values in this study could be attributed to the slightly larger number of participants (n=28 vs n=21) and possibly fewer outliers, which is supported by the excellent agreement. A further explanation for this finding may be that the Sydney system is commonly used by the assessors in this study.

There are a few factors that could have affected reliability of video assessment of the cervical branch. First, the cervical branch of the facial nerve innervates the platysma muscle, which tightens the anterior neck region and contributes to downward movement of the lower lip [44]. Downward movement of the lips has been found to have significantly more anterior-posterior axis than horizontal axis displacement, so the 2D nature of video assessment of the cervical branch could have affected reliability [27]. Furthermore, 2 assessors had noted poor visibility of the neck region during video assessment of patients who were wearing a hijab or jewelry around the neck. As the neck region is where the platysma muscle is situated, poor visibility of the area would have affected video assessment of the cervical branch. In a face-to-face assessment or tele-assessment over real-time videoconference, the assessor would be able to request the patient to remove any obstructions.

Consistently poor reliability for synkinesis assessment was observed between face-to-face and video assessments. As synkinesis is a multidimensional involuntary movement that occurs simultaneously in an area of the face different from the area being examined for voluntary movement (eg, during smiling, there could be synkinesis in the eye squinting) [26], it could be missed by the clinician who is directing attention to the area being examined for voluntary movement, which could happen in both face-to-face and tele-assessments. In video-based tele-assessment though, there is potential to pause and rewind footage to check for synkinesis.

The Bland-Altman plot indicated that patients could potentially be graded within 1 grade apart between face-to-face and video assessments of the temporal and buccal branches but more than 1 grade apart for that of the zygomatic, marginal mandibular, and cervical branches and synkinesis. This was not surprising as these movements (eg, forehead raise, nose wrinkle, smile, eye closure, and movement of the lower lips) produce a significant displacement in the anterior-posterior axis [27], which is difficult to adequately assess during a video assessment in the frontal plane.

In contrast, during a face-to-face assessment or tele-assessment over a videoconference media such as Skype, the clinician would have the opportunity to request a profile view of the patient, which would take into account movement occurring in the anterior-posterior axis. The clinician may also be able to request the removal of any obstruction for better visibility of the area being examined in real-time tele-assessment. Future research can build on current findings and compare the reliability of static video assessment with live videoconferencing assessment of facial paralysis.

Face-to-Face Versus Video-Assessed Sunnybrook System Scores

When implemented in face-to-face settings, reliability and agreement of the Sunnybrook system were similar to when used over a video. Reliability was poor to fair when assessing resting symmetry and synkinesis but good to excellent when assessing voluntary movement. Overall, agreement was generally good to excellent across all parameters. Similar to this study, previous research [26] has recorded good ICC reliability scores for the assessment of voluntary movement when using the Sunnybrook system; however, poor ICC reliability scores have been found for the assessment of synkinesis.

As synkinesis has been shown to be difficult to measure, clinicians may have asked patients to repeat and vary the speed and intensity of movements in the face-to-face setting to clarify their assessment finding. They, however, may not have chosen to rewind the video assessments when making ratings for this study. Furthermore, clinicians could not request a movement variation in terms of speed or intensity on the video, which may have unmasked the synkinesis as is sometimes done in face-to-face settings.

Second, studies on FNP have found patients' quality of life and emotional health to be affected to a larger extent by functional deficits caused by asymmetry in voluntary movements, such as the inability to smile symmetrically and oral incontinence, but to a smaller extent by synkinesis and resting symmetry [1,2,9,37]. Hence, assessors might have prioritized the rating of voluntary movement over resting symmetry and synkinesis as a result of their own personal experience in dealing with what patients perceive as their primary impairments. Third, the low ICC and fairly high PEA and PCA were indicators of outliers skewing the reliability measure or a restriction of range in the grades [40].

Again, reliability was similar and poor between face-to-face and video assessments. In addition, Kayhan et al [45] had found substantial reliability while using the Sunnybrook system to assess videos of patients, whereas Banks et al [31] had found excellent reliability when using the eFACE system to rate videos of participants. The poor reliability for synkinesis assessment may, therefore, be attributed to subjectivity in determining the degree of synkinesis among assessors instead of the 2D nature of video assessment.

In this study, the Sunnybrook composite score had excellent reliability in both face-to-face and video assessments. Hu et al [46], Ross and Nedzelski [35], and Kanerva et al [32] reported similar ICCs for the composite score, indicating excellent reliability. However, the components of the Sunnybrook system comprising resting symmetry, voluntary movement, and synkinesis would give more information about a patient's facial function during management of FNP than the composite score alone.

Bland-Altman plots indicated that video-assessed mean Sunnybrook scores for resting symmetry, voluntary movement, and synkinesis were similar to the face-to-face assessment scores. However, limits of agreement were too large when comparing video assessment of voluntary movement, synkinesis, and composite scores with face-to-face assessment, although they were reasonable for assessment of resting symmetry. This was expected as there is no displacement in the face at rest when no facial expression is performed, so most of what a clinician would be assessing would be in the vertical and horizontal axes of the frontal plane, and the lack of an anterior-posterior axis is not expected to make a difference in video assessment.

Overall, video-assessed Sunnybrook scores were generally as reliable as face-to-face assessed scores, but limits of agreement were too large between video and face-to-face assessments. This suggests that the Sunnybrook system could reasonably be implemented over a video, but video assessment would not necessarily substitute face-to-face assessment until there is further evidence proving reasonable limits of agreement between face-to-face and video assessments.

Limitations and Future Research

Although this exploratory study presents interesting findings about the potential for computer-mediated assessments of FNP, there are some limitations. First, the use of recorded videos, rather than real-time videoconferencing, may have limited ecological validity. In tele-health settings, assessment is interactive; patients can be asked to repeat or vary movements and to move the camera for closer or different views of particular movements. This is an opportunity for future research, to compare tele-assessment in real time, and to understand the impact of clinician-patient interactions on tele-assessment accuracy and reliability. It would also enable investigation into whether clinicians are likely to ask patients to repeat a movement despite not rewinding prerecorded videos of patients in this study.

Second, synkinesis was consistently poorly assessed as has been found in previous studies. To improve reliability of assessing synkinesis overall, regardless of the grading system used, protocols on assessing synkinesis could be improved and standardized, with additional notes about the need to review for synkinesis in the same area potentially multiple times. There could also be more specific and relevant training provided toward the protocols [26].

Third, the inter-rater reliability of expert assessors using facial nerve grading scales has been previously demonstrated; therefore, it was not evaluated in this study as each participant was evaluated by a different combination of assessors. A future study could potentially investigate this with a larger sample size.

Finally, the videos recorded in this study were of 1080p quality resolution, which is among the highest definition video modes available currently. Quality of the video assessment could potentially hinder reliability of data [22,23,25], which may affect observations, particularly if the changes in movement were subtle rather than very obvious.

Implications for Tele-Assessment of Facial Nerve Paralysis

This study suggests that further work is required for the 2D flat screen interface used in tele-health to be comparable with the face-to-face assessment of facial nerve disorders. The video assessment of FNP using the House-Brackmann facial grading scale and Sunnybrook facial grading system was generally as reliable as face-to-face assessment, though there was insufficient agreement between video and face-to-face assessments. This study also showed that the differences in reliability between face-to-face and video assessments using the Sydney system could likely be attributed to the lack of an anterior-posterior axis in the video assessment and generally poor reliability in assessment of synkinesis. Video assessment, however, does not take into account the opportunity for real-time clinician-patient interaction that would present in tele-assessment, which would allow clinicians to request patients to turn to their sides to show a profile view, to repeat certain movements, and to remove viewing obstructions such as face and neck coverings and jewelry, thereby potentially improving agreement between face-to-face and tele-assessments in all 3 facial grading systems.

As access to specialist care may pose an additional challenge for patients with FNP, future research could also compare the direct and indirect costs of telepractice with face-to-face management of FNP.

Conclusions

Our findings highlight the need for further research into the use of the House-Brackmann, Sydney, and Sunnybrook systems in tele-assessment of patients' facial function. Although reliability of face-to-face scores is similar to reliability of video-assessed scores, there is insufficient agreement between assessments in both modalities to strongly recommend the reliability of its use with current protocols, especially when synkinesis is present. This suggests that tele-assessment has future potential; however, research into the effect of fine-tuning facial movement protocols for use over the 2D screen to maximize reliability of a Web-based assessment is recommended. As evidence to support usage grows, the use of tele-health could increase access to specialized services for individuals with FNP, thereby improving quality of care and rehabilitation outcomes [12,18,19,23,25].

Conflicts of Interest

None declared.

References

1. Coulson SE, O'dwyer NJ, Adams RD, Croxson GR. Expression of emotion and quality of life after facial nerve paralysis. Otol Neurotol 2004 Nov;25(6):1014-1019. [Medline: 15547436]
2. Fu L, Bundy C, Sadiq SA. Psychological distress in people with disfigurement from facial palsy. Eye (Lond) 2011 Oct;25(10):1322-1326 [FREE Full text] [doi: 10.1038/eye.2011.158] [Medline: 21720412]
3. Eviston TJ, Croxson GR, Kennedy PG, Hadlock T, Krishnan AV. Bell's palsy: aetiology, clinical features and multidisciplinary care. J Neurol Neurosurg Psychiatry 2015 Dec;86(12):1356-1361. [doi: 10.1136/jnnp-2014-309563] [Medline: 25857657]
4. Evans RA, Harries ML, Baguley DM, Moffat DA. Reliability of the House and Brackmann grading system for facial palsy. J Laryngol Otol 1989 Nov;103(11):1045-1046. [Medline: 2607201]
5. Fattah AY, Gurusinghe AD, Gavilan J, Hadlock TA, Marcus JR, Marres H, Sir Charles Bell Society. Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. Plast Reconstr Surg 2015 Feb;135(2):569-579. [doi: 10.1097/PRS.0000000000000905] [Medline: 25357164]
6. Chong LS, Eviston TJ, Low TH, Hasmat S, Coulson SE, Clark JR. Validation of the clinician-graded electronic facial paralysis assessment. Plast Reconstr Surg 2017 Jul;140(1):159-167. [doi: 10.1097/PRS.0000000000003447] [Medline: 28338582]
7. Eren S, Dogan R, Ozturan O, Veyseller B, Hafiz A. How deleterious is facial nerve dissection for the facial nerve in parotid surgery: an electrophysiological evaluation. J Craniofac Surg 2017 Jan;28(1):56-60. [doi: 10.1097/SCS.0000000000003186] [Medline: 27893549]
8. Gaudin RA, Robinson M, Banks CA, Baiungo J, Jowett N, Hadlock TA. Emerging vs time-tested methods of facial grading among patients with facial paralysis. JAMA Facial Plast Surg 2016 Jul 01;18(4):251-257. [doi: 10.1001/jamafacial.2016.0025] [Medline: 27101446]
9. Volk GF, Granitzka T, Kreysa H, Klingner CM, Guntinas-Lichius O. Initial severity of motor and non-motor disabilities in patients with facial palsy: an assessment using patient-reported outcome measures. Eur Arch Otorhinolaryngol 2017 Jan;274(1):45-52. [doi: 10.1007/s00405-016-4018-1] [Medline: 27040558]
10. Helwig NE, Sohre NE, Ruprecht MR, Guy SJ, Lyford-Pike S. Dynamic properties of successful smiles. PLoS One 2017;12(6):e0179708 [FREE Full text] [doi: 10.1371/journal.pone.0179708] [Medline: 28658294]
11. Russell TG, Hoffmann TC, Nelson M, Thompson L, Vincent A. Internet-based physical assessment of people with Parkinson disease is accurate and reliable: a pilot study. J Rehabil Res Dev 2013;50(5):643-650 [FREE Full text] [Medline: 24013912]
12. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
13. Ward EC, Sharma S, Burns C, Theodoros D, Russell T. Managing patient factors in the assessment of swallowing via telerehabilitation. Int J Telemed Appl 2012;2012:132719 [FREE Full text] [doi: 10.1155/2012/132719] [Medline: 23008704]
14. Ward EC, Sharma S, Burns C, Theodoros D, Russell T. Validity of conducting clinical dysphagia assessments for patients with normal to mild cognitive impairment via telerehabilitation. Dysphagia 2012 Dec;27(4):460-472. [doi: 10.1007/s00455-011-9390-9] [Medline: 22271284]
15. Vranceanu AM, Riklin E, Merker VL, Macklin EA, Park ER, Plotkin SR. Mind-body therapy via videoconferencing in patients with neurofibromatosis: an RCT. Neurology 2016 Aug 23;87(8):806-814. [doi: 10.1212/WNL.0000000000003005] [Medline: 27449066]
16. Khoja S, Scott R, Husyin N, Durrani H, Arif M, Faqiri F, et al. Impact of simple conventional and telehealth solutions on improving mental health in Afghanistan. J Telemed Telecare 2016 Dec;22(8):495-498. [doi: 10.1177/1357633X16674631] [Medline: 27799454]

17. Luxton DD, Pruitt LD, Wagner A, Smolenski DJ, Jenkins-Guarnieri MA, Gahm G. Home-based telebehavioral health for US military personnel and veterans with depression: a randomized controlled trial. J Consult Clin Psychol 2016 Nov;84(11):923-934. [doi: 10.1037/ccp0000135] [Medline: 27599225]
18. Turkstra LS, Quinn-Padron M, Johnson JE, Workinger MS, Antoniotti N. In-person versus telehealth assessment of discourse ability in adults with traumatic brain injury. J Head Trauma Rehabil 2012;27(6):424-432 [FREE Full text] [doi: 10.1097/HTR.0b013e31823346fc] [Medline: 22190010]
19. Liss HJ, Glueckauf RI, Ecklund-Johnson EP. Research on telehealth and chronic medical conditions: critical review, key issues, and future directions. Rehabilit Psychol 2002;47(1):8-30. [doi: 10.1037//0090-5550.47.1.8]
20. Chavooshi B, Mohammadkhani P, Dolatshahi B. A randomized double-blind controlled trial comparing Davanloo Intensive Short-Term Dynamic Psychotherapy as internet-delivered vs treatment as usual for medically unexplained pain: a 6-month pilot study. Psychosomatics 2016;57(3):292-300. [doi: 10.1016/j.psym.2016.01.001] [Medline: 27017450]
21. Stasiak K, Fleming T, Lucassen MF, Shepherd MJ, Whittaker R, Merry SN. Computer-based and online therapy for depression and anxiety in children and adolescents. J Child Adolesc Psychopharmacol 2016 Apr;26(3):235-245. [doi: 10.1089/cap.2015.0029] [Medline: 26465266]
22. Lundell S, Holmner A, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015 Jan;109(1):11-26. [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
23. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
24. Engel CC, Jaycox LH, Freed MC, Bray RM, Brambilla D, Zatzick D, et al. Centrally assisted collaborative telecare for posttraumatic stress disorder and depression among military personnel attending primary care: a randomized clinical trial. JAMA Intern Med 2016 Jul 01;176(7):948-956. [doi: 10.1001/jamainternmed.2016.2402] [Medline: 27294447]
25. Demiris G, Oliver DP, Wittenberg-Lyles E, Washington K. Use of videophones to deliver a cognitive-behavioural therapy to hospice caregivers. J Telemed Telecare 2011;17(3):142-145 [FREE Full text] [doi: 10.1258/jtt.2010.100503] [Medline: 21303934]
26. Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. Otolaryngol Head Neck Surg 2005 Apr;132(4):543-549. [doi: 10.1016/j.otohns.2005.01.027] [Medline: 15806042]
27. Coulson SE, Croxson GR, Gilleard WL. Quantification of the three-dimensional displacement of normal facial movement. Ann Otol Rhinol Laryngol 2000 May;109(5):478-483. [doi: 10.1177/000348940010900507] [Medline: 10823477]
28. Berg T, Jonsson L, Engström M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. Otol Neurotol 2004 Nov;25(6):1020-1026. [Medline: 15547437]
29. Banks CA, Bhama PK, Park J, Hadlock CR, Hadlock TA. Clinician-graded electronic facial paralysis assessment: the eFACE. Plast Reconstr Surg 2015 Aug;136(2):223e-230e. [doi: 10.1097/PRS.0000000000001447] [Medline: 26218397]
30. Banks CA, Jowett N, Azizzadeh B, Beurskens C, Bhama P, Borschel G, et al. Worldwide testing of the eFACE facial nerve clinician-graded scale. Plast Reconstr Surg 2017 Feb;139(2):491e-498e. [doi: 10.1097/PRS.0000000000002954] [Medline: 28121888]
31. Banks CA, Jowett N, Hadlock TA. Test-retest reliability and agreement between in-person and video assessment of facial mimetic function using the eFACE Facial Grading System. JAMA Facial Plast Surg 2017 May 01;19(3):206-211 [FREE Full text] [doi: 10.1001/jamafacial.2016.1620] [Medline: 28006048]
32. Kanerva M, Poussa T, Pitkäranta A. Sunnybrook and House-Brackmann Facial Grading Systems: intrarater repeatability and interrater agreement. Otolaryngol Head Neck Surg 2006 Dec;135(6):865-871. [doi: 10.1016/j.otohns.2006.05.748] [Medline: 17141075]
33. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg 1985 Apr;93(2):146-147. [doi: 10.1177/019459988509300202] [Medline: 3921901]
34. Lee HY, Park MS, Byun JY, Chung JH, Na SY, Yeo SG. Agreement between the Facial Nerve Grading System 2.0 and the House-Brackmann Grading System in Patients with Bell Palsy. Clin Exp Otorhinolaryngol 2013 Sep;6(3):135-139 [FREE Full text] [doi: 10.3342/ceo.2013.6.3.135] [Medline: 24069515]
35. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg 1996 Mar;114(3):380-386. [doi: 10.1016/S0194-59989670206-1] [Medline: 8649870]
36. Croxson G, May M, Mester SJ. Grading facial nerve function: House-Brackmann versus Burres-Fisch methods. Am J Otol 1990 Jul;11(4):240-246. [Medline: 2399941]
37. Volk GF, Granitzka T, Kreysa H, Klingner CM, Guntinas-Lichius O. Nonmotor disabilities in patients with facial palsy measured by patient-reported outcome measures. Laryngoscope 2016 Dec;126(7):1516-1523. [doi: 10.1002/lary.25695] [Medline: 26421410]
38. Portney LG, Watkins MP. Statistical measures of reliability. In: Foundations of Clinical Research: Applications to Practice. 3rd edition. Pearson/Prentice Hall: Upper Saddle River, NJ; 2009:A-617.
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]

40. Rey JM, Plapp JM, Stewart G, Richards I, Bashir M. Reliability of the psychosocial axes of DSM-III in an adolescent population. Br J Psychiatry 1987 Feb;150:228-234 [FREE Full text] [Medline: 3651677]
41. MedCalc Software. Bland-Altman Plot URL: https://www.medcalc.org/manual/blandaltman.php [accessed 2018-10-27] [WebCite Cache ID 73U5ssTvO]
42. Myles P, Cui J. Using the Bland Altman method to measure agreement with repeated measures. British Journal of Anaesthesia 2007;99(3):309-311 [FREE Full text] [doi: 10.1093/bja/aem214] [Medline: 17702826 Version:1]
43. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
44. Eisele D, Smith RV. Complications of surgery of the neck. In: Complications in Head and Neck Surgery. 2nd edition. Philadelphia: Mosby; 2009.
45. Kayhan FT, Zurakowski D, Rauch SD. Toronto Facial Grading System: interobserver reliability. Otolaryngol Head Neck Surg 2000 Feb;122(2):212-215. [doi: 10.1016/S0194-5998(00)70241-5] [Medline: 10652392]
46. Hu WL, Ross B, Nedzelski J. Reliability of the Sunnybrook Facial Grading System by novice users. J Otolaryngol 2001 Aug;30(4):208-211. [Medline: 11771031]

Abbreviations

2D: 2-dimensional
3D: 3-dimensional
FNP: facial nerve paralysis
ICC: intraclass correlation coefficient
PCA: principal component analysis
PEA: percentage exact agreement
SFNC: Sydney Facial Nerve Clinic

Edited by M Focsa; submitted 28.05.18; peer-reviewed by S Paige, H Reychler; comments to author 18.06.18; revised version received 13.11.18; accepted 20.11.18; published 12.04.19.

Please cite as:
Tan JR, Coulson S, Keep M
Face-to-Face Versus Video Assessment of Facial Paralysis: Implications for Telemedicine
J Med Internet Res 2019;21(4):e11109
URL: https://www.jmir.org/2019/4/e11109/
doi:10.2196/11109
PMID:30977734

Original Paper

Midwives', Obstetricians', and Recently Delivered Mothers' Perceptions of Remote Monitoring for Prenatal Care: Retrospective Survey

Dorien Lanssens[1,2,3*], MSc, PhD; Thijs Vandenberk[1,2*], MSc, PhD; Joy Lodewijckx[1,2], MSc; Tessa Peeters[1,2], MSc; Valerie Storms[1,2], MSc, PhD; Inge M Thijs[1,2,3], MSc, PhD; Lars Grieten[1,2], Prof Dr; Wilfried Gyselaers[1,2,3], Prof Dr

[1]Limburg Clinical Research Program, Mobile Health Unit, Hasselt University, Diepenbeek, Belgium

[2]Faculty of Medicine and Life Sciences, Hasselt University, Diepenbeek, Belgium

[3]Department of Cardiology & Future Health, Ziekenhuis Oost-Limburg, Genk, Belgium

[*]these authors contributed equally

Corresponding Author:
Dorien Lanssens, MSc, PhD
Limburg Clinical Research Program
Mobile Health Unit
Hasselt University
Agoralaan
Diepenbeek, 3590
Belgium
Phone: 32 11 26 85 33
Email: dorien.lanssens@uhasselt.be

Abstract

Background: The Pregnancy Remote Monitoring (PREMOM) study enrolled pregnant women at increased risk of developing hypertensive disorders of pregnancy and investigated the effect of remote monitoring in addition to their prenatal follow-up.

Objective: The objective of this study was to investigate the perceptions and experiences of remote monitoring among mothers, midwives, and obstetricians who participated in the PREMOM study.

Methods: We developed specific questionnaires for the mothers, midwives, and obstetricians addressing 5 domains: (1) prior knowledge and experience of remote monitoring, (2) reactions to abnormal values, (3) privacy, (4) quality and patient safety, and (5) financial aspects. We also questioned the health care providers about which issues they considered important when implementing remote monitoring. We used a 5-point Likert scale to provide objective scores. It was possible to add free-text feedback at every question.

Results: A total of 91 participants completed the questionnaires. The mothers, midwives, and obstetricians reported positive experiences and perceptions of remote monitoring, although most of them had no or little prior experience with this technology. They supported a further rollout of remote monitoring in Belgium. Nearly three-quarters of the mothers (34/47, 72%) did not report any problems with taking the measurements at the required times. Almost half of the mothers (19/47, 40%) wanted to be contacted within 3 to 12 hours after abnormal measurement values, preferably by telephone.

Conclusions: Although most of midwives and obstetricians had no or very little experience with remote monitoring before enrolling in the PREMOM study, they reported, based on their one-year experience, that remote monitoring was an important component in the follow-up of high-risk pregnancies and would recommend it to their colleagues and pregnant patients.

Trial Registration: ClinicalTrials.gov NCT03246737; https://clinicaltrials.gov/ct2/show/NCT03246737 (Archived by WebCite at http://www.webcitation.org/76KVnHSYY)

(J Med Internet Res 2019;21(4):e10887) doi:10.2196/10887

KEYWORDS

remote monitoring; gestational hypertensive diseases; questionnaires; monitoring, ambulatory; hypertension, pregnancy-induced; surveys and questionnaires

Introduction

Background

Due to demographic changes and rapid improvements in medical technology, the health care sector is confronted with major challenges and great opportunities. The care and follow-up of a pregnant woman and her unborn baby is an important element in health care. Due to the changing lifestyles of pregnant women, the number of high-risk pregnancies has risen over the last few decades [1-3]. Therefore, there is a need to increase the efficiency of follow-up for these pregnancies without loss of quality of care. Telemedicine presents an opportunity for the follow-up of high-risk pregnancies.

Defined as the use of information and communication technologies for supporting health and health-related activities [4], telemedicine is not simply an addition to conventional care, but rather is implemented in current private and public health care approaches. Remote monitoring (RM) is a type of telemedicine that has a broad definition. It is useful for conducting medical practice from a distance and has been used in a wide variety of electronic health care applications [5]. RM can be performed either by live monitoring of vital parameters or asynchronously, whereby data obtained in the patient's home environment are sent to the health care provider [4]. Examples of chronic diseases that could benefit from RM include diabetes, heart failure, and cardiac arrhythmias [6-8].

The Pregnancy Remote Monitoring Study

The Pregnancy Remote Monitoring (PREMOM) study, which started in January 2015 in a tertiary center, Ziekenhuis Oost-Limburg (Genk, Belgium), involved RM of pregnant women at high risk of hypertensive disorders of pregnancy (HDP). The PREMOM study design, data collection method, and first promising results are described in detail elsewhere [9,10] (NCT03246737). Briefly, the PREMOM study was performed in the outpatient clinic of a second-level prenatal center where pregnant women with HDP received RM or conventional care. Women in the RM group received obstetric surveillance using a blood pressure monitor, an activity tracker, and a weight scale. They were asked to measure blood pressure twice a day, measure their weight once a week, and wear an activity tracker for 24 hours/day. These data were automatically sent by Wi-Fi or Bluetooth to an online platform, which was developed by the Mobile Health Unit (University of Hasselt, Hasselt, Belgium). A midwife reviewed the data every workday. The activity data were tracked to investigate the influence of daily activity (eg, total number of steps per day) on the development of HDP. Predetermined thresholds (systolic blood pressure >140 mm Hg, diastolic blood pressure >90 mm Hg, or weight gain >1 kg/day) were configured and resulted in automatically generated alarm signals.

The midwife discussed the alarm events with the obstetrician in charge to discuss the appropriate medical treatment. The midwife contacted the patients to give additional instructions about possible medical interventions such as altered medication regimens. These therapeutic interventions were according to local management.

Objectives

Because, to our knowledge, the perceptions or expectations of a prenatal RM follow-up program have not previously been investigated, we performed a quantitative survey of recently delivered women and health care providers (ie, both the obstetricians and the midwives). Here, we describe the main outcomes, which cover the following domains: (1) prior knowledge and experience of RM, (2) reactions to abnormal values, (3) privacy, (4) quality and patient safety, and (5) financial aspects. We also asked health care providers about important aspects to consider when implementing RM.

Methods

Questionnaires

The research group of the Mobile Health Unit designed 3 questionnaires: (1) for women who were followed up with RM during their last pregnancy, (2) for midwives working at the Ziekenhuis Oost-Limburg (Genk, Belgium) who were involved in the use of RM, and (3) for consulting obstetricians working at several hospitals in Limburg. The questionnaires assessed the 5 domains to elucidate PREMOM participants' perceptions and experiences of RM and were based on the 6 building blocks established by the Mobile Health Working Group of the Voka Health Community (Brussels, Belgium): (1) protection of data, privacy, and the use of big data; (2) national and international regulations and responsibility; (3) quality, accessibility, and patient safety; (4) technology and interoperability; (5) financial aspects and business models; and (6) supportive policy frameworks in telemedicine. Here we discuss the results of the descriptive PREMOM questionnaires on the domains prior knowledge and experience of RM, reactions to abnormal values, privacy, quality and patient safety, and financial aspects, which are important to health care providers for further implementation of RM. We drafted the questionnaires in April 2016 using Survey Monkey 2016 (SurveyMonkey Inc) for completion online. We assessed all questions using 5-point Likert scales to obtain objective scores. It was possible to add free-text feedback at every question.

Participants

We sent the questionnaires in April 2016 to the women, midwives, and obstetricians who participated in the PREMOM study in 2015. We excluded student midwives and doctors in training.

Data Collection

The study participants received an email from the research team with a link to the online survey. We sent email reminders to all participants at 9 and 23 days after the first invitation.

Analysis

We assessed mean scores and ranks for each question using descriptive analytical methods. The number of participants included in the analyses of individual questions was different from the total number of analyzed questionnaires because some mothers, midwives, and obstetricians did not complete all of the questions. We required at least half of the questionnaire to be completed for inclusion in the analysis. We conducted statistical analysis with IBM SPSS version 24.0 (IBM Corporation).

Ethical Considerations

We sent a generic link, to maintain anonymity, to the participants to fill in the survey. A bulk email was sent with the participants' email addresses included in the BCC field to ensure that there were no recognizable personal elements in the email.

The email was addressed to "Dear Madam" or "Dear Colleague" to remove the personal salutation to participate in this study. In addition, no personal participant identification number was requested or electronically reported when completing the questionnaires. Unique internet protocol addresses prevented duplicate responses to the questionnaires. The Medical Ethics Committee of Ziekenhuis Oost-Limburg approved this study (no. 14/078U).

Results

Participant Characteristics

The study population consisted of 158 people: 92 mothers (58%), 52 midwives (33%), and 14 obstetricians (9%). The total number of pregnant women involved in the PREMOM study was 119, so we contacted 77% (92/119) of the participants after their delivery. The 27 women who did not participate didn't answer their phone, didn't have an email address, or didn't speak Dutch. We excluded 1 obstetrician from the final analyses for completing less than 50% of the questionnaire. Therefore, the total response rate was 57% (91/158). Multimedia Appendix 1 shows the questionnaire and response summaries for midwives, Multimedia Appendix 2 shows the questionnaire and response summaries for obstetricians, and Multimedia Appendix 3 shows the questionnaire and response summaries for patients in English translation from the original Dutch. Table 1 shows the participants' demographic characteristics.

Prior Knowledge and Experience of Remote Monitoring

The first part of the questionnaire examined the midwives' and obstetricians' prior knowledge or experience of RM. Overall, 29 of the 35 midwives (83%) and 7 of the 9 (78%) obstetricians reported little or no experience of RM (Figure 1).

The midwives were also asked about their experience of RM as a threat to their daily work. The majority (29/35, 83%) of midwives did not perceive RM as a threat to their work.

Timing and Method of Communication in Case of an Event

Nearly three-quarters (34/47, 72%) of the participating mothers reported that they had no problems with taking the measurements at the requested times. Of the 7 mothers (15%) who reported difficulties with the recommended measurements, 4 (57%) were between 36 and 40 years old, 2 (29%) were between 26 and 30 years old, and 1 (14%) was between 31 and 35 years old.

We also asked participants about the acceptable time limit for being contacted by their health care provider in case of an unexpected event. Of the 47 women who completed the questionnaire, 13 (28%) preferred to be contacted within 3 hours of the event, 19 (40%) preferred to be contacted between 3 and 12 hours, and 15 (32%) preferred to be contacted more than 12 hours after the event (Table 2).

Table 1. Characteristics of participants.

Study group	Responses, n (%)
Women who were remotely monitored during their last pregnancy (n=47)	
Age (years)	
<20	0 (0)
20-25	5 (11)
26-30	16 (34)
31-35	21 (4)
36-40	4 (9)
>40	1 (2)
Primigravidity	
Primipara	21 (45)
Multipara	26 (55)
History of hypertensive disorders of pregnancy	
Yes	17 (36)
No	10 (21)
N/A[a]	20 (4)
Level of education	
Lower secondary school	4 (9)
Higher secondary school	12 (26)
High school	20 (43)
University	11 (23)
Midwives (n=35)	
Age (years)	
20-25	3 (9)
26-30	8 (23)
31-35	7 (20)
36-40	3 (9)
>40	14 (40)
Time in practice (years)	
<5	4 (11)
5-15	15 (43)
16-25	8 (23)
>25	8 (23)
Main activity on nursing unit	
Delivery unit	11 (31)
Maternity	8 (23)
Maternal intensive care	10 (29)
Prenatal visits	6 (17)
Obstetricians (n=9)	
Time in practice (years)	
<5	1 (11)
5-15	6 (67)
16-25	0 (0)

Study group	Responses, n (%)
>25	2 (22)
Main activity in their specialty	
Delivery unit	4 (44)
Obstetrics	4 (44)
Oncology	1 (11)

[a] N/A: not available.

Figure 1. Summary of responses from the midwives and obstetricians on the question "Please indicate with a score from 1 (strongly disagree) to 5 (strongly agree): I had already experience with RM before this study." RM: remote monitoring.

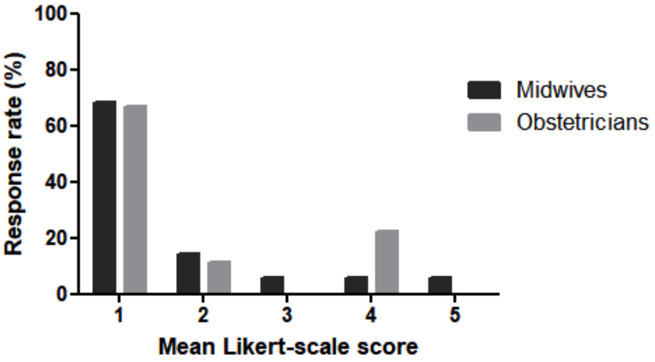


Table 2. Summary of responses to the question "Within how much time do you want to be contacted about events?" (n=47).

Response categories	Responses, n (%)
<3 hours	13 (28)
3-12 hours	19 (40)
12-24 hours	7 (15)
24-48 hours	5 (11)
>48 hours	3 (6)

Interestingly, 4 of the 5 mothers (80%) aged less than 25 years asked to be contacted within 3 hours of an event. We also asked the participants how to contact them following an event. The participants' first preference was to be contacted by telephone (weighted average 4.55/5), their second preference was during a prenatal consultation (weighted average 3.94/5), and the third preference was to be contacted by text messages (weighted average 3.17/5). Finally, we asked the participants who should contact the women in case of an event. The mothers and midwives stated that the obstetrician should be the first to contact the pregnant woman after an abnormal event. However, the obstetricians reported that their representing researcher should be the first health care provider to contact the pregnant woman in case of an event.

Privacy

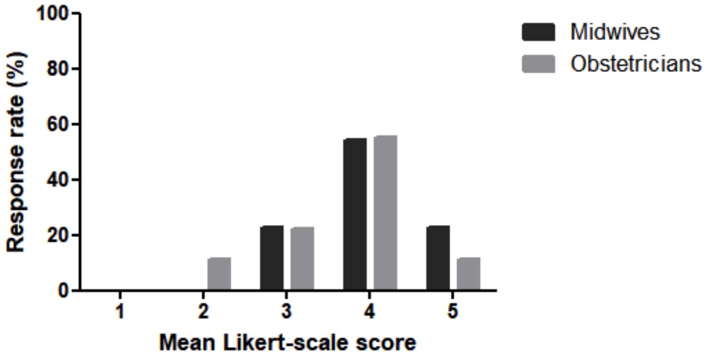
We asked the mothers if they felt that regularly sharing their health data was a threat to their privacy. Most (41/47, 87%) of the mothers reported that they did not have any negative concerns about privacy, while 3 mothers (aged 36-40 years) reported that sharing health data posed a threat to their privacy.

Quality and Patient Safety

We asked the mothers about the importance of RM in the follow-up of their pregnancy. Most (42/47, 89%) of the mothers had a positive response to this question. Meanwhile, 28 of the 35 (80%) midwives reported that RM provided added value to pregnant women, and 27 of the 35 (77%) midwives felt that RM improved the care for high-risk pregnancies. This percentage is slightly higher than that of the 9 obstetricians, 6 (67%) of whom felt that RM provided added value to their patients (Figure 2).

Moreover, 8 of the 9 (89%) obstetricians responded, based on their experience of the PREMOM study, that the pregnant women did not request additional prenatal consultations for the purpose of viewing their own vital parameters. Finally, 39 of the 47 (83%) mothers reported that RM gave them a feeling of safety.

Figure 2. Summary of responses from the midwives and obstetricians to the question "Do you believe that RM improves the care for pregnant women with an increased risk of gestational complications? Please indicate with a score from 1 (strongly disagree) to 5 (strongly agree)." RM: remote monitoring.



Financial Aspect

An important element in new health care practices is their financial cost. Therefore, the relative and absolute costs of each component in telemonitoring programs need to be evaluated. All 3 groups of participants reported that the cost of RM should be as low as possible, and about half of the mothers expected RM to be free, without a personal contribution from the patient (25/47, 53%). It is also important to obtain information on any potential payer of RM. The mothers expected the hospital to be the main payer, followed by their health insurance provider, whereas midwives and obstetricians felt that the pregnant women should also personally contribute to the cost of RM.

Further Implementation of Remote Monitoring

We asked the midwives and obstetricians about important factors to support the implementation of RM into daily practice. Most of the midwives (31/35, 89%) felt that it is important to receive additional training on "the information that must be given to pregnant women about hypertensive disorders of pregnancy and the added value of remote monitoring for this disease," as one of the midwives noted. Obstetricians (7/9, 78%) considered this 11 percentage points less necessary than did midwives. The obstetricians (8/9, 89%) felt that training on the technical handling of the devices (eg, installation and common problems) was the most important factor. About three-quarters of midwives (27/35, 77%) had the same response to this question. For the final evaluation of the project, we asked the obstetricians whether they would recommend RM to pregnant women and their colleagues. Overall, 6 of the 9 (67%) obstetricians supported this service and would recommend it to their patients, while 7 of the 9 (78%) obstetricians would recommend RM to their colleagues. Finally, 6 of the 9 (67%) obstetricians recommended that this follow-up should be expanded to all pregnant women in Belgium who are at increased risk of HDP.

Discussion

Principal Findings

RM is a relatively new field in obstetric research. Earlier studies of telemedicine that included cervical dilation and preterm labor as the main outcome demonstrated that transmitting uterine activity by telecommunication resulted in significantly prolonged pregnancy survivals [11,12]. Studies of telemedicine for patients with gestational diabetes mellitus demonstrated lower levels of frustration and concerns about their diabetes and a better acceptance of their diabetic condition [13], greater feelings of self-efficacy [14], and reduced unscheduled face-to-face visits [15,16] in the telemedicine group compared with the control group. Reduced costs [17,18] and greater feelings of maternal satisfaction [14,19,20] were obtained when telemedicine was used in obstetric care. Newborns had a higher gestational age at delivery [18] and were less likely to have a low birth weight [11,18] or to be admitted to the neonatal intensive care unit [11,18] in the telemedicine group compared with a control group. Fetuses with abnormal versus normal fetal heart rate at home monitoring were more likely to have a lower gestational age [21]. Recent studies about RM in women at risk for HDP demonstrated that they had fewer inductions, more spontaneous labors, and fewer maternal and neonatal hospitalizations when compared with conventional care [9,10]. Also, providing women at risk for HDP with RM was shown to be cost-effective for the health care system [22].

To our knowledge, this is the first quantitative survey of an RM program for prenatal care. The results show that most midwives and obstetricians had no or very little experience of RM before they participated in the PREMOM study. After taking part in the PREMOM study and the survey, the midwives reported that RM is not a threat to their daily work. Most of the mothers who were supervised by RM during their last pregnancy did not experience any problems with taking the required measurements at the specified times. Most of the mothers thought that it would be acceptable to be contacted within 3 to 12 hours after an abnormal value, and they preferred to be contacted by telephone.

The study of Giardina et al [23] showed the duality of feedback after a normal or an abnormal test. Nearly two-thirds of clinicians agreed that patients should receive direct feedback after a normal test. However, most physicians in the study expressed concerns about direct notification of clinically abnormal test results based on a patient's anxiety, confusion, lack of expertise to interpret the results, and seeking unreliable information to understand the results, and concerns that the patient would seek care without consulting their provider. The results of that study showed that doctors would be comfortable with a time interval of 24 to 48 hours for contacting a patient after an abnormal test result [23].

Privacy is a critical aspect of health care and RM [24]. The mothers in our study did not have concerns about sharing their health data with their obstetrician. As mentioned by Piwek and Ellis [25], data security and patients' privacy are essential elements for the adoption of digital smartphone research methods. Some risk-averse participants might be unwilling to share their clinical data with a commercial partner. However, none of the participants reported any privacy breaches using RM during this study.

The quality of care experienced by pregnant women with increased risk of HDP was enhanced by RM, as reported by the surveyed mothers and health care providers, and supported by the results of the prior pilot study [9]. Mothers who were involved in the project reported that RM gave them a feeling of security throughout their pregnancy. Previous research concluded that pregnant women with gestational diabetes mellitus had an increased sense of self-regulation when they used RM to send their blood glucose levels to their midwives [14,16]. Other research showed that pregnant women had heightened feelings of maternal satisfaction when using RM as additional care with their labor induction [19,20].

The mothers, midwives, and obstetricians included in this study reported that RM is an important aspect of the follow-up of high-risk pregnancies. An issue that raises important questions in telemedicine is the rather low adherence rate to RM, especially during long-term monitoring [26-29].

Measuring blood pressure, body weight, and activity every day is a prerequisite to ensure adequate monitoring of pregnant women, although this may appear burdensome to many of them. However, the mothers surveyed in this study did not experience this obstacle.

The obstetricians stated that they would recommend RM to colleagues and other pregnant women. Most of the obstetricians proposed extending RM to all women with high-risk pregnancies in Belgium. The obstetricians and midwives also reported that all users need additional training to support the implementation of RM. Earlier research already mentioned the challenge in training these obstetricians and midwives in the collection and interpretation of results, as well as incorporation of the remote patient data into routine clinical practice [30].

Strengths and Limitations of the Study

Despite the increased implementation of RM in health care, its use is still limited in obstetrics. To our knowledge, this was the first study to investigate obstetricians', midwives', and recently delivered mothers' perceptions of the use of RM for preterm follow-up of pregnancies at risk for HDP. Another strength of this study is that it included stakeholders involved in the use of RM, including health care providers and actual users. The questionnaires also allowed the participants to explain their responses to each question, allowing us to obtain supplementary information. Furthermore, the participants could complete the questionnaire anonymously. Finally, a relatively high percentage of participants in the PREMOM study completed the questionnaires.

Although the results of this study are encouraging, there are several limitations that should be considered for future research. First, because the questionnaire was completed anonymously, it was not possible to contact the individual participants to request additional information. Second, the questionnaire was digital and completed in an uncontrolled condition, so it is unclear whether the participants were exposed to external influences when they completed the questionnaire. Additionally, the 3 groups in this study had small sample sizes, which could affect external validity. Third, this study was performed in a local hospital, which can reduce the generalization of the results. Fourth, the study included obstetricians who worked at several hospitals in Limburg, but the midwives and mothers were enrolled from only a single center (Ziekenhuis Oost-Limburg).

Recommendations for Further Research

Both the mothers and the midwives felt that the obstetrician should be responsible for contacting the patient after an abnormal event, while the obstetricians suggested that their reporting researcher should be responsible for this task. This may relate to the organization of prenatal care in Belgium, where midwives act nearly as obstetric nurses rather than independently, and the prenatal care for pregnant women mostly is performed by an obstetrician, whether a pregnant woman has a high- or a low-risk pregnancy. It is remarkable that none of these 3 groups felt that this could be a task of the patient's midwife, although the researcher (DL) in this study is certified as a midwife. Still, the allocation of the responsibility for RM coordination to the midwives seems logical, as they act as an intermediary between the pregnant woman and the obstetrician. Clearly, further research is needed to understand the factors underlying this opinion and how it could be changed.

Additionally, both the mothers and the health care workers stated that RM should be offered for free or that they wanted to pay as little as possible for the RM services. Although we have conducted a cost-effectiveness study, which showed that RM makes saving costs possible for the health care system [22], we have not yet studied the willingness to pay. This study would have an additional value to set a price for RM services when the health care provider or the hospital requests it.

Further, although 67% of the obstetricians would recommend RM to their patients and 78% would recommend it to their colleagues, the obstetricians who would not recommend it did not give any reason for this. A follow-up qualitative questionnaire to investigate the underlying reasons for this should be helpful for the further implementation of RM in standard prenatal care for women at risk for HDP.

Interestingly, the mothers preferred to be contacted between 3 and 12 hours after an abnormal clinical measurement. This implies that the clinical data should be monitored 24 hours a day, 7 days a week in order to evaluate and interpret the vital parameters of pregnant women and permit an intervention if necessary. Therefore, we recommend developing a system of care aimed at providing these services. As our previous studies showed, our RM prenatal follow-up would result in the prenatal ward having a lower burden of treating women with HDP [9,10]. Finally, although we invited the mothers with abnormal events to additional prenatal consultations to assess fetal and maternal well-being, none of the patients or the participating obstetricians believed that this was needed and as such did not threaten to overload the health care system. These findings may contradict the statement that the medicalization of childbirth has gone too far and that too many medical interventions are performed in pregnancies, which has arisen from a variety of sources [31-36].

Conclusions

Although most midwives and obstetricians had no or very little experience with RM before they participated in the PREMOM study, they felt that it is an important aspect of the follow-up of pregnancies at risk for HDP. Most of the mothers who were supervised by RM during their last pregnancy thought that it was acceptable to be contacted within 3 to 12 hours after an abnormal value, and they preferred to be contacted by telephone. Most women had no concerns about regularly sharing their clinical data with their obstetrician, and they reported that RM gave them a feeling of security throughout their pregnancy. To our knowledge, this is the first quantitative survey of mothers, midwives, and obstetricians involved in an RM program in prenatal care. Further studies are needed to understand the underlying opinions of mothers, midwives, and obstetricians regarding RM. Based on our findings, we propose developing a care system with 24-hours-a-day, 7-days-a-week surveillance by RM of mothers at high risk of HDP.

Acknowledgments

This study was part of the Limburg Clinical Research Program UHasselt-Ziekenhuis Oost-Limburg-Jessa, supported by the foundation Limburg Sterk Merk, the province of Limburg, the Flemish government, Hasselt University, Ziekenhuis Oost-Limburg, and Jessa Hospital.

No specific funding was provided for the publication of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire for midwives, with responses.

[PDF File (Adobe PDF File), 274KB - jmir_v21i4e10887_app1.pdf]

Multimedia Appendix 2

Questionnaire for obstetricians, with responses.

[PDF File (Adobe PDF File), 313KB - jmir_v21i4e10887_app2.pdf]

Multimedia Appendix 3

Questionnaire for patients, with responses.

[PDF File (Adobe PDF File), 255KB - jmir_v21i4e10887_app3.pdf]

References

1. Forno E, Young OM, Kumar R, Simhan H, Celedón JC. Maternal obesity in pregnancy, gestational weight gain, and risk of childhood asthma. Pediatrics 2014 Aug;134(2):e535-e546 [FREE Full text] [doi: 10.1542/peds.2014-0439] [Medline: 25049351]
2. Nerenberg K, Daskalopoulou SS, Dasgupta K. Gestational diabetes and hypertensive disorders of pregnancy as vascular risk signals: an overview and grading of the evidence. Can J Cardiol 2014 Jul;30(7):765-773. [doi: 10.1016/j.cjca.2013.12.030] [Medline: 24726053]
3. Whitworth M, Dowswell T. Routine pre-pregnancy health promotion for improving pregnancy outcomes. Cochrane Database Syst Rev 2009 Oct 07(4):CD007536 [FREE Full text] [doi: 10.1002/14651858.CD007536.pub2] [Medline: 19821424]
4. Saner H, van der Velde E. eHealth in cardiovascular medicine: a clinical update. Eur J Prev Cardiol 2016 Oct;23(2 suppl):5-12. [doi: 10.1177/2047487316670256] [Medline: 27892420]
5. European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry. COCIR Telemedicine Toolkit for a better Deployment and Use of Telehealth. Brussels, Belgium: COCIR; 2011 May. URL: http://cocir.org/fileadmin/Publications_2011/telemedicine_toolkit_link2.pdf [accessed 2019-02-14] [WebCite Cache ID 76BDjfb2K]
6. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2009 Jan;32 Suppl 1:S62-S67 [FREE Full text] [doi: 10.2337/dc09-S062] [Medline: 19118289]
7. Ramani GV, Uber PA, Mehra MR. Chronic heart failure: contemporary diagnosis and management. Mayo Clin Proc 2010 Feb;85(2):180-195 [FREE Full text] [doi: 10.4065/mcp.2009.0494] [Medline: 20118395]
8. Levy S. Atrial fibrillation, the arrhythmia of the elderly, causes and associated conditions. Anadolu Kardiyol Derg 2002 Mar;2(1):55-60 [FREE Full text] [Medline: 12101796]

9. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
10. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
11. Corwin MJ, Mou SM, Sunderji SG, Gall S, How H, Patel V, et al. Multicenter randomized clinical trial of home uterine activity monitoring: pregnancy outcomes for all women randomized. Am J Obstet Gynecol 1996 Nov;175(5):1281-1285. [Medline: 8942501]
12. Wapner RJ, Cotton DB, Artal R, Librizzi RJ, Ross MG. A randomized multicenter trial assessing a home uterine activity monitoring device used in the absence of daily nursing contact. Am J Obstet Gynecol 1995 Mar;172(3):1026-1034. [Medline: 7892843]
13. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. J Telemed Telecare 2009;15(5):238-242. [doi: 10.1258/jtt.2009.081213] [Medline: 19590029]
14. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
15. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, de la Cruz MJ, Martín P, et al. A Telemedicine system based on Internet and short message service as a new approach in the follow-up of patients with gestational diabetes. Diabetes Res Clin Pract 2010 Feb;87(2):e15-e17. [doi: 10.1016/j.diabres.2009.12.002] [Medline: 20044162]
16. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
17. Buysse H, De Moor G, Van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. Int J Med Inform 2008 Jul;77(7):470-476. [doi: 10.1016/j.ijmedinf.2007.08.009] [Medline: 17923433]
18. Morrison J, Bergauer NK, Jacques D, Coleman SK, Stanziano GJ. Telemedicine: cost-effective management of high-risk pregnancy. Manag Care 2001 Nov;10(11):42-6, 48 [FREE Full text] [Medline: 11761593]
19. Rauf Z, O'Brien E, Stampalija T, Ilioniu FP, Lavender T, Alfirevic Z. Home labour induction with retrievable prostaglandin pessary and continuous telemetric trans-abdominal fetal ECG monitoring. PLoS One 2011;6(11):e28129 [FREE Full text] [doi: 10.1371/journal.pone.0028129] [Medline: 22140522]
20. O'Brien E, Rauf Z, Alfirevic Z, Lavender T. Women's experiences of outpatient induction of labour with remote continuous monitoring. Midwifery 2013 Apr;29(4):325-331. [doi: 10.1016/j.midw.2012.01.014] [Medline: 23159160]
21. Kuleva M, Salomon LJ, Benoist G, Ville Y, Dumez Y. The value of daily fetal heart rate home monitoring in addition to serial ultrasound examinations in pregnancies complicated by fetal gastroschisis. Prenat Diagn 2012 Aug;32(8):789-796. [doi: 10.1002/pd.3903] [Medline: 22653694]
22. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Vonck S, Claessens J, et al. Prenatal remote monitoring of women with gestational hypertensive diseases: cost analysis. J Med Internet Res 2018 Mar 26;20(3):e102 [FREE Full text] [doi: 10.2196/jmir.9552] [Medline: 29581094]
23. Giardina TD, Callen J, Georgiou A, Westbrook JI, Greisinger A, Esquivel A, et al. Releasing test results directly to patients: a multisite survey of physician perspectives. Patient Educ Couns 2015 Jun;98(6):788-796 [FREE Full text] [doi: 10.1016/j.pec.2015.02.011] [Medline: 25749024]
24. Beirlaen P, Bonroy B, Bruynseraede C, De Maesschalck L, De Maeyer C, De Raedt W, et al. [Mhealth as key to quality and affordability of care]. Whitepaper Voka Health Community mobile Health. Brussels, Belgium: Voka Health Community; 2015. URL: http://tinyurl.com/y2rmee9p [accessed 2019-02-14] [WebCite Cache ID 76BEu7lYg]
25. Piwek L, Ellis DA. Can programming frameworks bring smartphones into the mainstream of psychological science? Front Psychol 2016;7:1252 [FREE Full text] [doi: 10.3389/fpsyg.2016.01252] [Medline: 27602010]
26. Toogood PA, Abdel MP, Spear JA, Cook SM, Cook DJ, Taunton MJ. The monitoring of activity at home after total hip arthroplasty. Bone Joint J 2016 Nov;98-B(11):1450-1454. [doi: 10.1302/0301-620X.98B11.BJJ-2016-0194.R1] [Medline: 27803219]
27. Law LM, Edirisinghe N, Wason JM. Use of an embedded, micro-randomised trial to investigate non-compliance in telehealth interventions. Clin Trials 2016 Aug;13(4):417-424. [doi: 10.1177/1740774516637075] [Medline: 26968939]
28. Lyngå P, Fridlund B, Langius-Eklöf A, Bohm K. Perceptions of transmission of body weight and telemonitoring in patients with heart failure? Int J Qual Stud Health Well-being 2013;8:21524 [FREE Full text] [Medline: 24345687]
29. Jonassaint CR, Shah N, Jonassaint J, De Castro L. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]

30. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
31. Christiaens W, Van De Velde S, Bracke P. Pregnant women's fear of childbirth in midwife- and obstetrician-led care in Belgium and the Netherlands: test of the medicalization hypothesis. Women Health 2011 May;51(3):220-239. [doi: 10.1080/03630242.2011.560999] [Medline: 21547859]
32. Johanson R, Newburn M, Macfarlane A. Has the medicalisation of childbirth gone too far? BMJ 2002 Apr 13;324(7342):892-895 [FREE Full text] [Medline: 11950741]
33. Kuo SC. [A gender perspective on medicalized childbirth]. Hu Li Za Zhi 2015 Feb;62(1):10-15. [Medline: 25631179]
34. Parry DC. “We wanted a birth experience, not a medical experience”: exploring Canadian women’s use of midwifery. Health Care Women Int 2008 Sep;29(8):784-806. [doi: 10.1080/07399330802269451] [Medline: 18726792]
35. Sedigh Mobarakabadi S, Mirzaei Najmabadi K, Ghazi Tabatabaie M. Ambivalence towards childbirth in a medicalized context: a qualitative inquiry among Iranian mothers. Iran Red Crescent Med J 2015 Mar;17(3):e24262 [FREE Full text] [doi: 10.5812/ircmj.24262] [Medline: 26019905]
36. Shaw JCA. The medicalization of birth and midwifery as resistance. Health Care Women Int 2013;34(6):522-536. [doi: 10.1080/07399332.2012.736569] [Medline: 23514572]

Abbreviations

HDP: hypertensive disorders of pregnancy
PREMOM: Pregnancy Remote Monitoring
RM: remote monitoring

Edited by G Eysenbach; submitted 26.04.18; peer-reviewed by D Oepkes, YK Akpak, Y Motoki; comments to author 07.10.18; revised version received 02.12.18; accepted 31.12.18; published 15.04.19.

Please cite as:
Lanssens D, Vandenberk T, Lodewijckx J, Peeters T, Storms V, Thijs IM, Grieten L, Gyselaers W
Midwives’, Obstetricians’, and Recently Delivered Mothers’ Perceptions of Remote Monitoring for Prenatal Care: Retrospective Survey
J Med Internet Res 2019;21(4):e10887
URL: https://www.jmir.org/2019/4/e10887/
doi:10.2196/10887
PMID:30985286

Original Paper

Clinical Efficacy of Telemedicine Compared to Face-to-Face Clinic Visits for Smoking Cessation: Multicenter Open-Label Randomized Controlled Noninferiority Trial

Akihiro Nomura[1,2,3], MD, PhD; Tomoyuki Tanigawa[1,4], MD; Tomoyasu Muto[5], MSc; Takafumi Oga[6], MD; Yasushi Fukushima[7], MD; Arihiro Kiyosue[8], MD; Masaki Miyazaki[9], MD; Eisuke Hida[10], PhD; Kohta Satake[1,5], MPH, MBA, MD

[1]CureApp Institute, Karuizawa, Japan
[2]Innovative Clinical Research Center, Kanazawa University, Kanazawa, Japan
[3]Department of Cardiology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan
[4]Graduate School of Public Health, St. Luke's International University, Tokyo, Japan
[5]CureApp Inc, Tokyo, Japan
[6]Shinjuku Research Park Clinic, Tokyo, Japan
[7]Fukuwa Clinic, Tokyo, Japan
[8]Tokyo-eki Center Building Clinic, Tokyo, Japan
[9]Miyazaki Respiratory Care Clinic, Tokyo, Japan
[10]Department of Biostatistics and Data Science, Osaka University Graduate School of Medicine, Osaka, Japan

Corresponding Author:
Akihiro Nomura, MD, PhD
Innovative Clinical Research Center
Kanazawa University
13-1 Takaramachi
Kanazawa, 920-8641
Japan
Phone: 81 76 265 2049
Email: anomura@med.kanazawa-u.ac.jp

Abstract

Background: Tobacco is a major public health concern. A 12-week standard smoking cessation program is available in Japan; however, it requires face-to-face clinic visits, which has been one of the key obstacles to completing the program, leading to a low smoking cessation success rate. Telemedicine using internet-based video counseling instead of regular clinic visits could address this obstacle.

Objective: This study aimed to evaluate the efficacy and feasibility of an internet-based remote smoking cessation support program compared with the standard face-to-face clinical visit program among patients with nicotine dependence.

Methods: This study was a randomized, controlled, open-label, multicenter, noninferiority trial. We recruited nicotine-dependent adults from March to June 2018. Participants randomized to the telemedicine arm received internet-based video counseling, whereas control participants received standard face-to-face clinic visits at each time point in the smoking cessation program. Both arms received a CureApp Smoking Cessation smartphone app with a mobile exhaled carbon monoxide checker. The primary outcome was a continuous abstinence rate (CAR) from weeks 9 to 12. Full analysis set was used for data analysis.

Results: We randomized 115 participants with nicotine dependence: 58 were allocated to the telemedicine (internet-based video counseling) arm and 57, to the control (standard face-to-face clinical visit) arm. We analyzed all 115 participants for the primary outcome. Both telemedicine and control groups had similar CARs from weeks 9 to 12 (81.0% vs 78.9%; absolute difference, 2.1%; 95% CI –12.8 to 17.0), and the lower limit of the difference between groups (–12.8%) was greater than the prespecified limit (–15%).

Conclusions: The application of telemedicine using internet-based video counseling as a smoking cessation program had a similar CAR from weeks 9 to 12 as that of the standard face-to-face clinical visit program. The efficacy of the telemedicine-based smoking cessation program was not inferior to that of the standard visit–based smoking cessation program.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry: UMIN000031620; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000035975.

(J Med Internet Res 2019;21(4):e13520) doi:10.2196/13520

KEYWORDS

smoking cessation; nicotine dependence; telecare; telemedicine; mHealth; digital therapeutics; mobile phone

Introduction

Tobacco is a major public health concern and the biggest preventable cause of a variety of disorders such as cerebro- and cardiovascular diseases, malignant tumors, and chronic obstructive pulmonary disease [1,2]. In Japan, the estimated number of smokers is more than 20 million, and smoking is responsible for approximately 130,000 deaths per year [3]. Thus, reducing the prevalence of smoking would help prevent deaths from life-threatening diseases [4].

To help quit smoking, Japan provides a smoking cessation program for patients with nicotine dependence. This 12-week program mainly consists of face-to-face clinic visits, involving counseling with a primary physician, checking exhaled carbon monoxide (CO) concentration, and prescribing smoking cessation medications [5]. Face-to-face clinic visits enable physicians to directly perform counseling, physical examinations, and various tests in person. However, despite these intensive efforts to have patients complete the program, more than half of the program participants could not complete the entire program [6]. The majority of patients with nicotine dependence in Japan are typically employed men who are extremely busy and unwilling to spend a large portion of their day visiting the clinic to receive the smoking cessation program [6]. Program dropout is normally considered equivalent to smoking cessation failure [7]. Therefore, promoting dedication and completion of the smoking cessation program could be crucial to making these people succeed at quitting smoking.

Recently, telemedicine, defined as remote delivery of health care via the internet, was considered one of the useful methods for providing medical care to patients [8,9]. Telemedicine minimizes patients’ burden of visiting a health institution and waiting for consultations with their physicians. Telemedicine could also be suitable for delivering a smoking cessation support program. Considering the preliminary report that 75% of participants could complete the smoking cessation program when conducted via telemedicine, the requirement of regular face-to-face visits at a clinic might be the reason for the low completion rate of the smoking cessation program [10]. Therefore, telemedicine that enables the smoking cessation program participants to receive their regular counseling via the internet could have the potential to improve the overall smoking cessation success rate among patients with nicotine dependence by providing them easier access to the program. However, it remains uncertain whether telemedicine using an internet-based video counseling system is effective for delivering the smoking cessation program compared to the standard face-to-face clinical visits.

In this study, we tested the clinical efficacy and feasibility of telemedicine using internet-based Web counseling compared to a standard face-to-face clinical visit in the smoking cessation program among patients with nicotine dependence.

Methods

Trial Design and Participants

This trial was a randomized, controlled, open-label, multicenter, noninferiority trial. Details of the trial protocol have been described elsewhere [11]. In brief, participants in both arms underwent the smoking cessation program used in Japan [5]. For the telemedicine arm, the entire smoking cessation program was conducted remotely via an internet-based video counseling system, except for the first registration visit [12]. For the control arm, participants followed the standard smoking cessation program conducted through face-to-face clinic visits. Participants in both arms used the CureApp Smoking Cessation (CASC) system [13]. The primary outcome was a biochemically validated continuous abstinence rate (CAR) from weeks 9-12.

We recruited individuals with nicotine dependence from March to June 2018. We conducted follow-up for 24 weeks. Only the participants who met all the inclusion criteria were included; those who met any of the exclusion criteria were excluded [11]. Briefly, we included participants who were diagnosed with nicotine dependence (Tobacco Dependence Screener score ≥5 points) [14], had a Brinkman index ≥200, had the will to quit smoking immediately, agreed to undergo the smoking cessation program, and could use a smartphone. We excluded participants who had severe mental illness, could not tolerate the follow-up for 6 months, had used smoking cessation supplements or medication before the registration, planned to use any smoking cessation aids or to participate in any kind of smoking, or had regular clinic visits for diseases other than nicotine dependence planned within 12 weeks of registration.

Primary physicians at each clinic obtained written informed consent from all trial participants. We confirmed that clinics participating in this trial could provide the standard smoking cessation support program and had the necessary equipment to provide Web-based telemedicine (eg, WiFi access in the facility). We conducted this trial in compliance with the Declaration of Helsinki, Medical Device Good Clinical Practice guidelines, and all other applicable laws and guidelines in Japan. The trial protocol was approved by the Tokyo-Eki Center-Building Clinic institutional review board. We reported the trial according to CONSORT-EHEALTH (V 1.6.1). This trial was registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN000031620).

Randomization

We used the stratified-block randomization (four blocks) method with a 1:1 allocation ratio to achieve equal assignment to two arms with stratification of the trial sites. Participants were allocated to either the telemedicine arm or the control arm. The randomization was performed by the staff at each participating clinic at the time of participants' registration, using a computer-generated random sequence.

Procedures

Participants allocated to the telemedicine arm received internet-based Web counseling for the smoking cessation program. Participants assigned to the control arm received the conventional face-to-face clinic visits for the smoking cessation program. Both arms also received the CASC smartphone app and a mobile exhaled CO checker during the trial period (24 weeks).

The standard smoking cessation program in Japan consists of five face-to-face clinic visits lasting for 12 weeks, including doctor consultations and exhaled CO checks at a registered institution or clinic [5]. All study participants visited their primary physicians at their first visit to confirm that they fully understood the trial protocol. At this visit, the physicians decided to prescribe appropriate smoking cessation medication, to provide guidance in accordance with the standard program procedure, and to provide participants with the CASC smartphone app integrated with a mobile CO checker.

Following the first visit, telemedicine participants were supposed to receive counseling via the internet-based video counseling system with a standardized telemedicine platform application [12] instead of visiting their clinics to see their primary physicians. As in the standard program, they met with their physicians via video counseling at each planned visit at weeks 2, 4, 8, 12, and 24. The control participants were supposed to visit their clinics at weeks 2, 4, 8, 12, and 24.

Overview of CureApp Smoking Cessation

The CASC system was developed by CureApp, Inc. (Tokyo, Japan). Details of the system have been demonstrated elsewhere [13]. The CASC system consists of the CASC smartphone app [15], mobile exhaled CO checker, and Web-based personal computer for primary physicians. For the telemedicine arm, the primary physicians provided the app prescription code to the participants at their first visit in the outpatient clinics. The telemedicine participants downloaded the app through their smartphones; activated the app by entering the code; and keyed in their baseline data, motivation, and self-confidence regarding smoking cessation. The CASC smartphone app has four main components: (1) keeping a smoking cessation digital diary (filled in once a day), (2) lectures and educational videos helping its users to quit smoking, (3) interactive counseling by chat-bot, and (4) daily measurement and recording of exhaled CO concentration levels at home using the mobile CO checker. The Web-based personal computer software for the primary physicians provided a data-management app from patients' CASC smartphone apps and advice for physicians to follow the national clinical guidelines.

Outcomes

The primary outcome was the biochemically validated CAR from weeks 9-12, which was consistent with the previous clinical trial of varenicline in Japan [16]. This measure is defined as the percentage of individuals continuously not smoking (success) during the specified period. This study defined smoking cessation success as self-reported continuous abstinence as well as exhaled CO concentration ≤ 10 ppm during the given period [16]. For example, CAR from weeks 9-12 indicates that the patient achieved smoking cessation success if he/she self-reported continuous abstinence for 9-12 weeks and his/her exhaled CO concentration was ≤10 ppm at week 12. We also evaluated the following secondary outcomes: CAR from weeks 9-24; changes in the scores on the Mood and Physical Symptoms Scale (MPSS) [17] and the 12-item French version of the Tobacco Craving Questionnaire (FTCQ-12) [18]; the Kano Test for Social Nicotine Dependence (KTSND) score [19] at weeks 8, 12, and 24; Nicotine Dependence Cognition Scale (NDCS) score at weeks 12 and 24 [11]; and all adverse events during the trial.

Sample Size and Inferiority Margin

From previous pilot studies of a CASC smartphone app, the difference in CAR from 9-12 weeks between the CASC smartphone app group (78%) and historical control groups (not using the app; 54%) is 24% [20-22]. Therefore, we hypothesized that the telemedicine group would not provide clinically worse CAR from 9-12 weeks compared to the control group (prespecified margin of 15% based on estimated 80% CARs in both groups) [11]. We calculated the required sample size as 114 (57 per each arm) based on the precision of estimate that the lower limit of the 95% CI of the difference between treatment effects exceeded the threshold amount of 15%. Therefore, we aimed to recruit at least 114 participants to allow for this sample size.

Statistical Analysis

We compared all endpoints between the telemedicine and control groups. Baseline characteristics were described by means and SDs, medians and interquartile ranges (for continuous variables), or proportions (for categorical variables). We analyzed the primary outcome using the full analysis set (excluding participants who violated the inclusion or exclusion criteria). In case a patient discontinued an allocated treatment, the case was considered as smoking cessation failure. We compared CARs between telemedicine and the control groups using a logistic regression model with crude odds ratios (ORs). For all outcomes, summary statistics and group difference measures (eg, ORs by logistic regression or mean differences) were presented with 95% CIs. We also tested if the CAR from 9-12 weeks of telemedicine and the control condition varied by subgroups. We assessed interactions of CAR from 9-12 weeks with each variable (greater than vs less than the median for continuous variables) and calculated ORs with 95% CIs in each subgroup. We used the Wilcoxon signed-rank test for comparing the scores regarding nicotine dependence between baseline and at weeks 12 and 24 in each group. R version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria) was used for all the analyses.

Results

We randomized 115 participants to the telemedicine arm or control arm (Figure 1). Each arm was well balanced in the baseline characteristics (Table 1). All participants were prescribed smoking cessation medication at registration: varenicline for 55% and the nicotine patch for 45% of the participants. During the trial period, three participants discontinued their allocated treatment, two participants by consent withdrawal and one participant by loss to follow-up. Finally, all 115 participants were enrolled for further analyses.

Biochemically validated CARs from weeks 9-12 were 81.0% (95% CI 71-91) in the telemedicine group and 78.9% (95% CI 68-89) in the control group (Table 2). The absolute difference was 2.1% (95% CI –12.8 to 17.0); the lower limit of the 95% CI (–12.8%) was greater than the prespecified limit of –15% (Multimedia Appendix 1). The OR was 1.14 (95% CI 0.45-2.88). Moreover, CARs from weeks 9-24 were 74.1% (95% CI 63-85) in the telemedicine group and 71.9% (95% CI 60-84) in the control group. Therefore, there were no statistically significant differences between the telemedicine and control groups in CARs. In terms of program feasibility, adherence rates by session were both high (over 95%) during the trial (Multimedia Appendix 1). In addition, there were no serious or device-related adverse events in the groups during the trial (Multimedia Appendix 1). CASC product issues during the trial were reported by four participants (3.5%): two issues were related to mobile CO checker connection failure (one in telemedicine group and one in control group), one was related to login failure in the control group, and one was related to other malfunction of the primary physician interface (failure of a chart closure) in the telemedicine group.

Next, we demonstrated the efficacy of the telemedicine on CAR from weeks 9-12 by subgroup. Although the KTSND had a moderate interaction with the outcome, we found little evidence of significant interactions for any of the subgroup analyses (Multimedia Appendix 1).

We also assessed the evolution of scores by MPSS, FTCQ-12, KTSND, and NDCS (Multimedia Appendix 1). During the 24-week trial period, all FTCQ-12, KTSND, and NDCS scores were significantly decreased over time in both groups. MPSS scores, especially regarding urges (“time spent with urges” and “strength of urges”), were also significantly reduced over time in both groups.

Figure 1. Trial flowchart. CASC: CureApp Smoking Cessation; SC: smoking cessation.

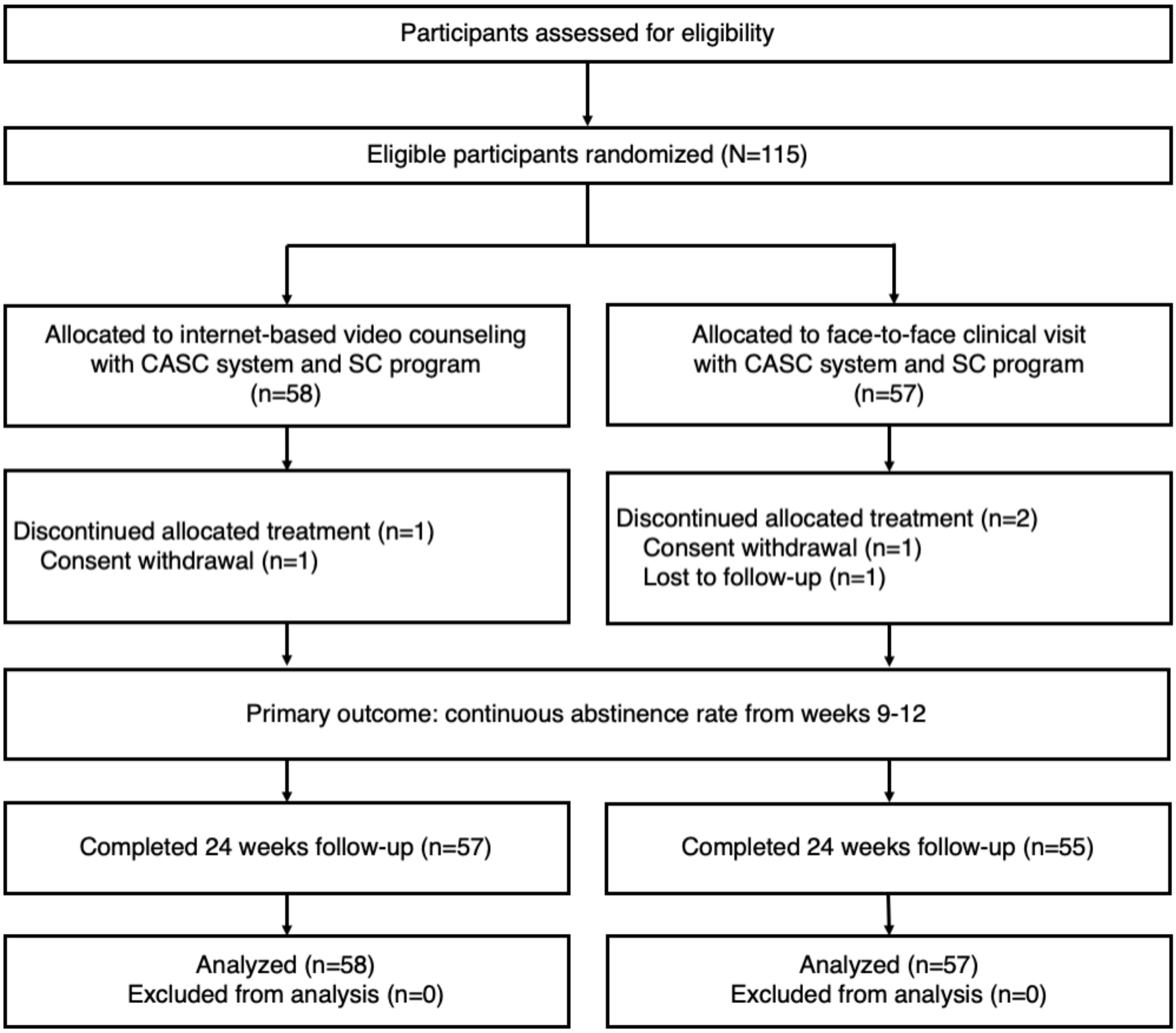


Table 1. Baseline characteristics of the trial participants.

Characteristic	Total (N=115)	Telemedicine (N=58)	Standard care (N=57)
Age (years), mean (SD)	55 (11)	55 (12)	53 (10)
Male sex, n (%)	93 (81)	45 (78)	48 (84%)
Body mass index (kg/m^2), median (interquartile range)	23 (21-26)	23 (21-26)	23 (21-25)
Brinkman index, median (interquartile range)	**480 (325-720)**	**500 (340-755)**	**450 (310-720)**
Cigarettes per day	15 (13-20)	16 (15-20)	15 (12-20)
Years of smoking	34 (27-40)	35 (29-41)	33 (26-38)
Number of attempts of smoking cessation before the trial, median (interquartile range)	1 (0-2)	1 (0-2)	0 (0-2)
TDS[a] score, median (interquartile range)	7 (6-8)	7 (6-8)	7 (6-8)
FTND[b] score, median (interquartile range)	5 (3-7)	5 (4-7)	5 (3-6)
KTSND[c] score, median (interquartile range)	17 (16-20)	17 (16-20)	18 (15-21)
Comorbidities, n (%)			
Hypertension	33 (29)	16 (28)	17 (30)
Diabetes mellitus	9 (8)	5 (9)	4 (7)
Dyslipidemia	42 (37)	23 (40)	19 (33)
Medication, n (%)			
Varenicline	63 (55)	29 (50)	34 (60)
Nicotine patch	52 (45)	29 (50)	23 (40)

[a] TDS: Tobacco Dependence Screener.

[b] FTND: Fagerström Test for Nicotine Dependence.

[c] KTSND: Kano Test for Social Nicotine Dependence.

Table 2. Continuous abstinence rates in percentages from weeks 9-12 (primary outcome) and weeks 9-24 (secondary outcome).

Continuous abstinence rate	Telemedicine, mean (SE)	Control, mean (SE)	Difference (95% CI)	Odds ratio (95% CI)
Weeks 9-12	81.0 (5.1)	78.9 (5.4)	2.1 (–12.8 to 17.0)	1.14 (0.45-2.88)
Weeks 9-24	74.1 (5.7)	71.9 (6.0)	2.2 (–14.0 to 18.4)	1.12 (0.49-2.57)

Discussion

Principal Results

In this randomized trial, we assessed the efficacy and feasibility of an internet-based remote smoking cessation program compared to the standard face-to-face clinical visit program among patients with nicotine dependence. We found that (1) CARs from weeks 9-12 were relatively high in both groups (81.0% in the telemedicine and 78.9% in the face-to-face control groups) and (2) the clinical efficacy of the telemedicine group was not significantly greater than that of the face-to-face control group in terms of CAR from weeks 9-12.

Comparison with Prior Work

This trial has several important findings. First, CARs from weeks 9-12 were relatively high. CARs from weeks 9-24 also reached favorable results in both groups (74.1% in telemedicine and 71.9% in control). Compared to other countries [23-25], these CARs estimated using the Japanese smoking cessation program were higher. One of the main reasons could be that all participants in the Japanese smoking cessation program need to swear to quit smoking and sign a declaration of smoking cessation before participating in the program. This is a unique feature of the Japanese smoking cessation program, and it would strongly contribute to selecting highly motivated participants who want to quit smoking. In terms of previous reports from Japan, Nakamura et al reported CARs from weeks 9-12 and weeks 9-24 of 65.4% and 37.7%, respectively, in nicotine-dependent patients receiving 1 mg varenicline on prescription [16]. Japan's Ministry of Health, Labour and Welfare also reported the latest success rate of 63.8% for smoking cessation at week 12 in nicotine-dependent patients undergoing the standard smoking cessation program [6]. Our trial showed comparable results with these studies, in terms of smoking cessation success rates. In addition, the efficacy of the telemedicine program was not inferior to that of the face-to-face clinic visit program over a prespecified limit value. In summary, telemedicine, or online-based video counseling, could be a viable alternative for managing a smoking cessation program, as it has shown a noninferior efficacy compared to the standard face-to-face clinic visit smoking cessation program.

Second, the dropout rates of the smoking cessation program were low even at week 24 in both groups, indicating potential improvement in the overall smoking cessation success rates. The Japanese national survey on the efficacy of nicotine-dependence treatment showed a linear relationship between the number of patient visits to outpatient clinics and the treatment success rate [6]. We used the CASC system for both groups in this trial. The system could cover intervals between counseling sessions or clinic visits that help patients obtain clinical guidance. It could also continuously monitor, promote, and encourage commitment to the smoking cessation program [13]. Thus, the CASC system, including the CASC smartphone app, might indirectly contribute to preventing dropouts from both groups.

Third, most of the scores for nicotine dependence in the trial were significantly decreased in both groups. This result indicated that the telemedicine-based and face-to-face smoking cessation programs with the CASC system were effective in improving the status of nicotine dependence and ameliorating craving for smoking over time in the smoking cessation program. However, mood symptoms (depressed, irritable in control, restless, hungry, and poor concentration) of the MPSS did not change during the trial. This outcome may be because the baseline median scores of these symptoms were already low (1=Not at all or 2=Slightly) and continuously stabilized during the trial period.

Strengths and Limitations

The strength of this trial was that it was the first randomized controlled trial to test the efficacy and feasibility of telemedicine using internet-based video counseling directly compared with face-to-face clinical visit in patients with nicotine dependence. This trial had a few limitations. First, the prespecified limit value of 15% might not be conservative. However, CARs from weeks 9-12 in both groups were almost what we expected (both around 80%), and we considered 65% of CAR from weeks 9-12 to be a reasonable threshold for checking the clinical relevance of the telemedicine program compared with the face-to-face clinic visit program in accordance with previous reports [6,16]. Second, this Web- and smartphone-based program might not be applied for individuals who cannot buy or access mobile devices and did not have enough literacy to read or access the mobile technology. Third, concluding the efficacy of the telemedicine in this 3-months trial could be difficult. Further trials lasting longer than 3 months might be needed to confirm the long-term efficacy of telemedicine. Fourth, we did not collect data on adherence of smoking cessation medications that might affect the trial results.

Conclusions

Telemedicine using internet-based video counseling for the smoking cessation program had a similar CAR from weeks 9-12 as that of the standard face-to-face clinic visit program. The efficacy of the telemedicine-based smoking cessation program was noninferior to that of the standard clinic visit–based smoking cessation program. The results of this trial demonstrated that internet-based counselling might be a viable alternative to standard clinic visits for smoking cessation.

Acknowledgments

We thank all the staff and participants involved in this trial. We would also like to express our gratitude to Ayaka Kato for her productive comments on our draft manuscript and Yoshihiko Raita and Masaaki Yamada for their support in conducting the statistical analyses. In addition, we thank Tatsue Takami, Satomi Ono, Naomi Suzuki, Masaru Tukahara and all other administrative and engineering team members in CureApp Inc for their dedicated work for this study.

Conflicts of Interest

This trial was supported by CureApp, Inc. AN received consulting fees from CureApp, Inc. TT and TM are employees of CureApp, Inc. KS is the founder and a shareholder of CureApp, Inc, and patent holder of the CASC system. EH has a consultation contract as a biostatistician with CureApp, Inc. TO, YF, AK and MM received research support fees from CureApp, Inc.

Multimedia Appendix 1

Supplemental figures and tables.

[PDF File (Adobe PDF File), 269KB - jmir_v21i4e13520_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 3MB - jmir_v21i4e13520_app2.pdf]

References

1. Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med 2012 Jan;9(1):e1001160 [FREE Full text] [doi: 10.1371/journal.pmed.1001160] [Medline: 22291576]

2. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2017 URL: http://apps.who.int/iris/bitstream/handle/10665/255874/9789241512824-eng.pdf [accessed 2019-04-19] [WebCite Cache ID 77k2EXTmH]
3. Ikeda N, Saito E, Kondo N, Inoue M, Ikeda S, Satoh T, et al. What has made the population of Japan healthy? Lancet 2011 Sep 17;378(9796):1094-1105. [doi: 10.1016/S0140-6736(11)61055-6] [Medline: 21885105]
4. Igarashi A, Goto R, Suwa K, Yoshikawa R, Ward AJ, Moller J. Cost-Effectiveness Analysis of Smoking Cessation Interventions in Japan Using a Discrete-Event Simulation. Appl Health Econ Health Policy 2016 Feb;14(1):77-87 [FREE Full text] [doi: 10.1007/s40258-015-0204-3] [Medline: 26597111]
5. The Standard Procedure Book for Smoking Cessation, 6th edition (in Japanese). 2014 Apr. URL: https://www.jrs.or.jp/uploads/uploads/files/information/non-smoking_06.pdf [accessed 2018-12-10] [WebCite Cache ID 74ZVOvjoU]
6. Ministry of Health, Labour and Welfare (Japan). 2017. Report on the Efficacy of Smoking Cessation Treatment Based on Nicotine Dependency Management Fee (in Japanese) URL: http://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000192293.pdf [accessed 2018-04-29] [WebCite Cache ID 74U5NcjPx]
7. Caponnetto P, Polosa R. Common predictors of smoking cessation in clinical practice. Respir Med 2008 Aug;102(8):1182-1192. [doi: 10.1016/j.rmed.2008.02.017] [Medline: 18586479]
8. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
9. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
10. Kitada M. Evaluation on the impact of telemedicine for clinic visit for treatment of nicotine dependence and its treatment success. Kinen-Kagaku 2017;11(15):1.
11. Tanigawa T, Nomura A, Kuroda M, Muto T, Hida E, Satake K. A Multi-center, Open-label, Randomized Controlled Trial Comparing Telemedicine and Face-to-Face Consultation Based on the Standard Smoking Cessation Program for Nicotine Dependence: Study Protocol. JMIR Res Protoc (upcoming) 2019:1. [doi: 10.2196/preprints.12701]
12. Ministry of Health, Labour and Welfare (Japan). 2018. Guideline for Online Clinical Practice (in Japanese) URL: https://www.mhlw.go.jp/file/05-Shingikai-10801000-Iseikyoku-Soumuka/0000201789.pdf [accessed 2019-04-19] [WebCite Cache ID 77k3M4hHr]
13. Nomura A, Tateno H, Masaki K, Muto T, Suzuki S, Satake K, et al. A Novel Smoking Cessation Smartphone App Integrated With a Mobile Carbon Monoxide Checker for Smoking Cessation Treatment: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Feb 11;8(2):e12252 [FREE Full text] [doi: 10.2196/12252] [Medline: 30741641]
14. Kawakami N, Takatsuka N, Inaba S, Shimizu H. Development of a screening questionnaire for tobacco/nicotine dependence according to ICD-10, DSM-III-R, and DSM-IV. Addict Behav 1999;24(2):155-166. [Medline: 10336098]
15. Masaki K, Tateno H, Kameyama N, Morino E, Watanabe R, Sekine K, et al. Impact of a Novel Smartphone App (CureApp Smoking Cessation) on Nicotine Dependence: Prospective Single-Arm Interventional Pilot Study. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12694. [doi: 10.2196/12694]
16. Nakamura M, Oshima A, Fujimoto Y, Maruyama N, Ishibashi T, Reeves KR. Efficacy and tolerability of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, in a 12-week, randomized, placebo-controlled, dose-response study with 40-week follow-up for smoking cessation in Japanese smokers. Clin Ther 2007 Jun;29(6):1040-1056. [doi: 10.1016/j.clinthera.2007.06.012] [Medline: 17692720]
17. West R, Hajek P. Evaluation of the mood and physical symptoms scale (MPSS) to assess cigarette withdrawal. Psychopharmacology (Berl) 2004 Dec;177(1-2):195-199. [doi: 10.1007/s00213-004-1923-6] [Medline: 15179542]
18. Berlin I, Singleton EG, Heishman SJ. Validity of the 12-item French version of the Tobacco Craving Questionnaire in treatment-seeking smokers. Nicotine Tob Res 2010 May;12(5):500-507 [FREE Full text] [doi: 10.1093/ntr/ntq039] [Medline: 20335281]
19. Otani T, Yoshii C, Kano M, Kitada M, Inagaki K, Kurioka N, et al. Validity and reliability of Kano Test for Social Nicotine Dependence. Ann Epidemiol 2009 Nov;19(11):815-822. [doi: 10.1016/j.annepidem.2009.03.016] [Medline: 19457683]
20. Masaki K, Tateno H, Morino E. A Multi-center, Single-Arm Study to Evaluate Efficacy of Smartphone Application - CureApp - for Smoking Cessation: Interim Report. 2017 Jul 09 Presented at: The 634th Kanto Regional Meeting, Japanese Society of Internal Medicine; Sept 07, 2017; Tokyo, Japan.
21. Ministry of Health, Labour and Welfare (Japan). 2008. Report on the Survey of Success Rate of Smoking Cessation Treatment at Health Insurance-Covered Medical Facilities to Reimburse Nicotine Dependency Management Fee (in Japanese) URL: http://www.mhlw.go.jp/shingi/2008/07/dl/s0709-8k.pdf [accessed 2019-11-04] [WebCite Cache ID 74U5umcPo]
22. Masaki K, Tateno H, Kameyama N. A Multi-center, Single-Arm Study to Evaluate Efficacy of Smartphone Application - CureApp - for Smoking Cessation. 2018 Apr 14 Presented at: The Japanese Society of Internal Medicine, Annual Scientific Meeting 2018; Apr 14, 2018; Kyoto, Japan.
23. Gonzales D, Rennard SI, Nides M, Oncken C, Azoulay S, Billing CB, Varenicline Phase 3 Study Group. Varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs sustained-release bupropion and placebo for smoking cessation: a randomized controlled trial. JAMA 2006 Jul 05;296(1):47-55. [doi: 10.1001/jama.296.1.47] [Medline: 16820546]

24. Jorenby D, Hays J, Rigotti N, Azoulay S, Watsky E, Williams K, Varenicline Phase 3 Study Group. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: a randomized controlled trial. JAMA 2006 Jul 05;296(1):56-63. [doi: 10.1001/jama.296.1.56] [Medline: 16820547]
25. Tonstad S, Tønnesen P, Hajek P, Williams KE, Billing CB, Reeves KR, Varenicline Phase 3 Study Group. Effect of maintenance therapy with varenicline on smoking cessation: a randomized controlled trial. JAMA 2006 Jul 05;296(1):64-71. [doi: 10.1001/jama.296.1.64] [Medline: 16820548]

Abbreviations

CAR: continuous abstinence rate
CO: carbon monoxide
FTCQ-12: 12-item French version of the Tobacco Craving Questionnaire
KTSND: Kano Test for Social Nicotine Dependence
MPSS: Mood and Physical Symptoms Scale
NDCS: Nicotine Dependence Cognition Scale
OR: odds ratio

Edited by G Eysenbach; submitted 28.01.19; peer-reviewed by N Gordon, C Mavrot, M Mahoney, J Murray, N Nguyen, A Al-Zalabani; comments to author 02.03.19; revised version received 26.03.19; accepted 11.04.19; published 26.04.19.

Please cite as:
Nomura A, Tanigawa T, Muto T, Oga T, Fukushima Y, Kiyosue A, Miyazaki M, Hida E, Satake K
Clinical Efficacy of Telemedicine Compared to Face-to-Face Clinic Visits for Smoking Cessation: Multicenter Open-Label Randomized Controlled Noninferiority Trial
J Med Internet Res 2019;21(4):e13520
URL: http://www.jmir.org/2019/4/e13520/
doi:10.2196/13520
PMID:30982776

Original Paper

Identification of Anxiety and Depression Symptoms in Patients With Cancer: Comparison Between Short and Long Web-Based Questionnaires

Susanne Mattsson[1], RN, PhD; Erik Martin Gustaf Olsson[2], PhD; Maria Carlsson[1], RN, PhD; Birgitta Beda Kristina Johansson[3], RN, PhD

[1]Lifestyle and Rehabilitation in long term illness, Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden
[2]Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[3]Experimental and Clinical Oncology, Department of immunology, genetics and pathology, Uppsala University, Uppsala, Sweden

Corresponding Author:
Susanne Mattsson, RN, PhD
Lifestyle and Rehabilitation in long term illness
Public Health and Caring Sciences
Uppsala University
Box 564
Uppsala, 75122
Sweden
Phone: 46 184710000 ext 6637
Fax: 46 184716675
Email: susanne.mattsson@pubcare.uu.se

Abstract

Background: Physicians and nurses in cancer care easily fail to detect symptoms of psychological distress because of barriers such as lack of time, training on screening methods, and knowledge about how to diagnose anxiety and depression. National guidelines in several countries recommend routine screening for emotional distress in patients with cancer, but in many clinics, this is not implemented. By inventing screening methods that are time-efficient, such as digitalized and automatized screenings with short instruments, we can alleviate the burden on patients and staff.

Objective: The aim of this study was to compare Web-based versions of the ultrashort electronic Visual Analogue Scale (eVAS) anxiety and eVAS depression and the short Hospital Anxiety and Depression Scale (HADS) with Web-based versions of the longer Montgomery Åsberg Depression Rating Scale-Self-report (MADRS-S) and the State Trait Anxiety Inventory- State (STAI-S) with regard to their ability to identify symptoms of anxiety and depression in patients with cancer.

Methods: Data were obtained from a consecutive sample of patients with newly diagnosed (<6 months) breast, prostate, or colorectal cancer or with recurrence of colorectal cancer (N=558). The patients were recruited at 4 hospitals in Sweden between April 2013 and September 2015, as part of an intervention study administered via the internet. All questionnaires were completed on the Web at the baseline assessment in the intervention study.

Results: The ultrashort and short Web-based-delivered eVAS anxiety, eVAS depression and HADS were found to have an excellent ability to discriminate between persons with and without clinical levels of symptoms of anxiety and depression compared with recommended cutoffs of the longer instruments MADRS-S and STAI-S (area under the curve: 0.88-0.94). Cutoffs of >6 on HADS anxiety and >7 hundredths (hs) on eVAS anxiety identified patients with anxiety symptoms with high accuracy. For HADS depression, at a cutoff of >5 and eVAS depression at a cutoff of >7 hs, the accuracy was very high likewise.

Conclusions: The use of the short and ultrashort tools, eVAS and HADS, may be a suitable initial method of Web-based screening in busy clinical settings. However, there are still a proportion of patients who lack access to the internet or the ability to use it. There is a need to find solutions for this group to find all the patients with psychological distress.

(J Med Internet Res 2019;21(4):e11387) doi:10.2196/11387

KEYWORDS

screening; cancer; depression; anxiety; internet; eHealth

Introduction

Symptoms of anxiety and depression are common in people diagnosed with cancer depending on the specific diagnosis, stage of the disease, sex, age, and the employed screening instrument [1,2]. According to a large Canadian study (n=10,153), almost 1 of 5 patients experienced clinical anxiety symptoms and clinical depression symptoms were present in about 12.87% (1204/9357) of the patients [1]. We found that 35.6% (176/495) of patients recently referred to the oncology department (<1 month) because of cancer disease experienced anxiety or depression symptoms and that more than 1 of 5 experienced such symptoms 6 months later [3]. Anxiety and depression in patients with cancer are associated with poor health-related quality of life, disease-related morbidity, poor treatment adherence, and prognosis [3-8].

Physicians and nurses in cancer care easily fail to detect symptoms of psychological distress because of lack of time and knowledge about how to diagnose anxiety and depression [9]. Consequently, many patients have an unmet need for psychosocial support [10]. Therefore, national guidelines in several countries recommend routine screening for emotional distress in patients with cancer [8,11-13]. On the contrary, routine clinical screening is still controversial and partly questioned because studies have yielded ambiguous results [14-16]. Our previous research project concluded that systematic screening for anxiety and depression is feasible in a clinical oncological setting and increases referral for clinical assessment, psychosocial support, and treatment [3]. However, the implementation of screening in routine care, without additional research funding, places significant challenges on an already strained cancer care. Lack of resources and absence of screening strategies, as well as receipt of appropriate aftercare, have been defined as barriers for a successful implementation of screening [2,9].

Screening for anxiety and depression using validated questionnaires is a standard procedure in research. Commonly used questionnaires vary considerably with regard to the number of items [16]. Health professionals prefer shorter questionnaires as they are less time-consuming than longer ones [9]. Ultrashort questionnaires (1 to 4 items) may be as successful as short questionnaires (5 to 20 items), hence efficient and acceptable, and are therefore suggested to be suitable for initial assessment in clinical settings [17]. Long questionnaires generally have a higher specificity, but as the clinical acceptance is low because of the higher demand on patients and staff, they are not always useful in routine cancer care [18]. Findings from research projects evaluating Web-based screening followed by psychosocial support, when distress is identified, indicate that this may be an efficient way to reduce suffering [19-21]. Thus, Web-based screening of anxiety and depression with automatic calculations of scores can be a realistic alternative to paper-based screening in a busy cancer clinic and may be one way to increase availability and referral for support and treatment when needed. An important first step is to evaluate the accuracy of Web-based versions of clinically relevant questionnaires for screening of anxiety and depression symptoms.

Digital versions of self-report symptom scales have shown high interformat reliability compared with pen and paper versions [22]. However, the researchers concluded that these findings could not be generalized to all questionnaires and settings and that future studies regarding the evaluation of Web-based versions of traditional paper-based questionnaires should include large sample sizes. In addition, further evaluations are needed because of the novelty of the procedure of Web-based screening.

In this study, we aimed to evaluate the accuracy of the Web-based versions of the ultrashort electronic Visual Analogue Scale (eVAS) and the short Hospital Anxiety and Depression Scale (HADS) [23,24] using the Web-based versions of the longer Montgomery-Åsberg Depression Rating Scale-Self-assessment (MADRS-S) [25] and the State-Trait Anxiety Inventory-State (STAI-S) [26] as references.

Methods

Participants and Procedures

A consecutive sample of patients with newly diagnosed (<6 months) breast, prostate, or colorectal cancer or with recurrence of colorectal cancer were recruited at 4 hospitals in Sweden between April 2013 and September 2015, as part of an intervention study administered via the internet [27]. The exclusion criteria were inability to read and understand Swedish, cognitive disability (eg, dementia or psychosis), a constant need for hospital care (Karnofsky score <40) or short expected survival (<3 months). The participants were approached and informed by a research assistant at a regular visit to the clinic and gave their written informed consent. This study was approved by the Regional Ethical Review Board in Uppsala (Dnr. 2012/003).

Data Collection

Quality Register

Background data regarding the disease and treatment were obtained from the diagnosis-specific quality registers of the Uppsala–Örebro region in Central Sweden.

Questionnaires

All questionnaires were completed on the Web at the baseline assessment in the intervention study [27], in a place of the participant's choice. Demographic information (age, sex, marital status, and education) was obtained via project-specific questions.

eVAS measures anxiety and depression on 2 scales ranging from 1 to 100 hundredths (hs), that is, independent of the screen size. The paper-based VAS is evaluated to perform well when correlated to standardized measures of anxiety and depression [23]. We chose separate scales for anxiety and depression instead of a single distress scale (eg, the Distress Thermometer) [28] for a more accurate comparison with MADRS-S and STAI-S. The patients were asked to grade their levels of anxiety and depression by making a mark somewhere between 0 (no anxiety or depression) and 100 (extreme anxiety or depression) on each scale. Web-based screening with the similar ultrashort Distress Thermometer has been successfully used in a large sample of cancer patients [29].

The HADS consists of 7 questions measuring depression and 7 measuring anxiety [24] and is the most validated and widely used questionnaire in screening for anxiety and depression in patients with cancer [30]. The patients were asked to rate their emotional status during the past week on a scale of 0 to 3. The total score of each scale is 21 points. A higher score implies more symptoms, 8 to 10 points indicate doubtful cases and 11 to 21 points indicate clinically significant cases of anxiety or depression, according to Zigmond and Snaith [24]. However, further evaluations of the HADS have revealed large variations in recommended cutoff scores for depression and anxiety in various groups of patients with cancer, both higher and lower scores have been suggested as cutoffs for clinically significant symptoms [31-33]. Thus, the optimal cutoff scores for identification of patients with a need for support are still not decided. Screening with computerized versions of the HADS in patients with cancer has been deemed feasible [34]. In addition, a comparison between a Web-based-delivered version and a paper version of HADS has yielded comparable findings with regard to frequencies and intensity of symptoms [35].

The STAI-S was used as a reference to eVAS and HADS anxiety. STAI-S is a self-administered questionnaire that measures participants' state of anxiety [26]. It has been used in patients with various cancer diagnoses and administered on the Web [36]. STAI-S comprises 20 items where the respondents are asked to rate their current feelings on a 4-point scale, from *not at all* (1) *to very much* (4). The scores are summed, ranging from 20 to 80 [26]. A cutoff level of >39 was used to indicate clinically significant symptoms of anxiety in this study based on previous research in similar populations [37,38].

The MADRS-S, the self-rating version of the MADRS [25], was used as a reference to eVAS and HADS depression. The MADRS-S has acceptable psychometric properties with regard to reliability, validity, and sensitivity to change [39]. In addition, it can be transferred to the Web without affecting the psychometric properties in a clinically significant way [40]. The MADRS-S has 9 items [41], and the respondents are asked to rate their emotional status during the previous 3 days. All items range from 0 to 6, and the total score is 54. A higher value indicates more depressive symptoms. The cutoff score for depression varies greatly in studies, but the thresholds recommended by Svanborg and Ekselius (0-12=minimal, 13-19=mild, 20-34=moderate, and >34=severe) [41] have been used in previous Swedish studies where the instrument was administered both on the Web and in paper and pen versions [42-44]. A cutoff of >12 has been applied in this study to indicate clinically significant symptoms of depression [41].

Data Analysis

The statistical analyses were performed using IBM Statistical Package for the Social Sciences (version 20.0) [45]. Receiver operating curve (ROC) analyses were used to assess the sensitivity and specificity of eVAS and HADS using STAI-S and MADRS-S as references. The ROCs depict the tradeoff between specificity and sensitivity for every possible cutoff score on the HADS and eVAS indexes. The area under the curve (AUC) can be interpreted as the probability that a randomly selected patient with anxiety or depression according to the STAI-S or MADRS-S, respectively, will score higher on the respective HADS and eVAS scales than a randomly selected patient without anxiety or depression according to the STAI-S or MADRS-S, respectively [46]. By calculating the AUC, it was possible to estimate the overall discriminative performance of the HADS and the eVAS with regard to the identification of patients with clinically significant levels of anxiety or depression symptoms. An AUC of 0.50 indicates a level of accuracy no better than chance and 0.70 to 0.80 represents acceptable discrimination. An AUC of 0.80 represents excellent discrimination, and an AUC of 1.0 indicates a test with perfect accuracy.

The specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to establish relevant cutoff scores for each of the ultrashort and short instruments. Sensitivity, specificity, and NPV are the most important indicators for establishing cutoffs as the aim of the screening is to detect as many as possible and to make sure that as few as possible are false negatives.

Results

Participants

A total of 1748 patients were assessed for eligibility, of whom 251 were excluded and 771 declined participation mainly because of a lack of internet access or lack of interest. A total of 726 (48.5%, 726/1497) study-eligible patients were enrolled and 558 (76.9%, 558/726) of them completed the questionnaires and constituted the final sample. A majority were under treatment, for example, chemotherapy, radiotherapy, or endocrine therapy. Mean age was lower in the final study sample than in the group not included. There was also a preponderance of patients with prostate cancer and a significantly smaller group of patients with colorectal cancer among the respondents compared with not included patients (Table 1).

Table 1. Demographic and clinical characteristics of the patients with cancer.

Characteristics	Respondents (N=558)	Not included[a] (N=939)	*P* value
Sex, n (%)			
Male	247 (44.3)	364 (38.8)	.14
Female	311 (55.7)	575 (61.2)	.14
Age (years), mean (range)	61 (29-86)	69 (24-99)	<.001
Diagnosis, n (%)			
Breast cancer	272 (48.7)	469 (49.9)	.001
Colorectal cancer	96 (17.2)	240 (25.6)	.001
Prostate cancer	190 (34.1)	230 (24.5)	.001
Marital status, n (%)			
Married or cohabiting	435 (78.0)	—[b]	—
Living alone	24 (4.3)	—	—
Living apart	86 (15.4)	—	—
Missing[c]	13 (2.3)	—	—
Education, n (%)			
Elementary school	114 (20.4)	—	—
High school	173(31.0)	—	—
University-level	262 (47.0)	—	—
Missing	9 (1.6)	—	—
Breast cancer, n (%)			
In situ	5 (1.8)	—	—
T0 No obvious primary tumor	14 (5.1)	—	—
T1-T4	246 (90.5)	—	—
≥1 lymph node involved	39 (14.3)	—	—
Distant metastases	4 (1.5)	—	—
Surgery	258 (94.9)	—	—
Radiation therapy	231 (84.9)	—	—
Chemotherapy	121 (44.5)	—	—
Endocrine therapy	191 (70.2)	—	—
Missing[c]	7 (2.6)	—	—
Prostate cancer, n (%)			
Unassessable tumor (TX)	2 (1.1)	—	—
T1-T4	183 (96.3)	—	—
≥1 lymph node involved	15 (7.9)	—	—
Distant metastases	15 (7.9)	—	—
Surgery	34 (17.9)	—	—
Radiation therapy	86 (45.3)	—	—
Chemotherapy	2 (1.1)	—	—
Expectancy or surveillance	26 (13.7)	—	—
Endocrine therapy	30 (15.8)	—	—
Missing[c]	5 (2.6)	—	—
Colorectal cancer, n (%)			

Characteristics	Respondents (N=558)	Not included[a] (N=939)	*P* value
Primary tumor not found (T0)	6 (6.3)	—	—
T1-T4	70 (72.9)	—	—
≥1 lymph node involved	34 (35.4)	—	—
Distant metastases	5 (5.2)	—	—
Surgery	77 (80.2)	—	—
Radiation therapy	22 (22.9)	—	—
Chemotherapy	34 (35.4)	—	—
Targeted drugs	39 (40.6)	—	—
Missing[c]	20 (20.8)	—	—

[a] Not included: internal dropouts (n=168) and declined participation (n=771).

[b] Data not available.

[c] Missing tumor stage.

Cases of Anxiety and Depression

The STAI-S (>39) identified 165 (30.0%, 165/550) cases and 385 noncases of anxiety, and MADRS-S (>12) identified 107 (19.4%, 107/551) cases and 444 noncases of depression. HADS anxiety identified 126 (22.6%, 126/558) participants with at least mild levels of anxiety and HADS depression identified 79 (14.2%, 79/558) participants with at least mild symptoms of depression, both scales using a cutoff score of >7. The vast majority of the scores on eVAS anxiety and depression were within the range of 1 to 20 hs (81.3%, 451/555 and 82.8%, 457/552, respectively; Table 2).

Anxiety—Accuracy of Hospital Anxiety and Depression Scale and Electronic Visual Analogue Scale Compared With State Trait Anxiety Inventory

Using STAI-S as the reference instrument and >39 as the cutoff for clinical symptoms of anxiety, the AUC for HADS anxiety was 0.93 (95% CI 0.91-0.95). Both 5 and 6 were cutoffs with good sensitivity and specificity. A score >5 identified 95% of the patients with symptoms and ruled out 71% of those without and >6 identified 88% of the patients with symptoms and ruled out 81% of those without. The corresponding AUC for the eVAS anxiety was 0.90 (95% CI 0.87-0.93), and a cutoff score of >7 hs identified 86% of the patients with symptoms of anxiety and ruled out 83% of the patients without. The NPV values, that is, the proportion of all the undetected cases that did not have symptoms according to the reference, were generally high for the mentioned cutoffs (≥90%), whereas the PPV values, that is, the proportion of all the detected cases that actually had symptoms according to the reference, were somewhat lower (≥67%). These most important values correspond to an excellent overall accuracy (Table 3).

Depression—Accuracy of Hospital Anxiety and Depression Scale and Electronic Visual Analogue Scale Compared With Montgomery Åsberg Depression Rating Scale-Self-Report

Using MADRS-S as the reference instrument and >12 as the cutoff for clinical symptoms of depression, the AUC for HADS depression was 0.94 (95% CI 0.92-0.96). Both 5 and 6 were cutoffs with good sensitivity and specificity. A score >5 identified 91% of the patients with symptoms and ruled out 83% of those without, and a score >6 identified 87% of those with symptoms and ruled out 89% of those without. The corresponding AUC for the eVAS depression was 0.88 (95% CI 0.84-0.93). A cutoff score of >7 hs is a cutoff with good sensitivity and specificity. It identified 85% of the patients with depressive symptoms and ruled out 83% of those without. The NPVs were generally high for the mentioned cutoffs (≥94%) whereas the PPVs were lower (≥56%), indicating a relatively large portion of false positives among the detected. With the exception of PPV, these values also correspond to an excellent overall accuracy (Table 3).

Table 2. Cases and noncases of anxiety and depression symptoms in patients with newly diagnosed breast, colorectal, or prostate cancer.

Measures	Statistics			Cronbach alpha
	n (%)	Mean (SD)	Range	
State Trait Anxiety Inventory-State (n=550)	—[a]	34.8 (11.1)	20-73	.95
Non-cases (<40)	385 (70.0)	—	—	—
Cases (>39)	165 (30.0)	—	—	—
HADS[b] A (n=558)	—	5 (8.7)	0-18	.88
0-7	432 (77.4)	—	—	—
8-10	73 (13.1)	—	—	—
11-21	53 (9.5)	—	—	—
eVAS[c] A (n=555)	—	12.2 (18.8)	—	—
0-20	451 (81.3)	—	—	—
21-40	48 (8.6)	—	—	—
41-60	29 (5.2)	—	—	—
61-80	24 (4.3)	—	—	—
81-100	3 (0.6)	—	—	—
Montgomery Åsberg Depression Rating Scale-Self-report (n=551)	—	7 (7)	0-37	.88
Non-cases (<13)	444 (80.6)	—	—	—
Cases (>12)	107 (19.4)	—	—	—
HADS D (n=558)	—	3.7 (3.4)	0-18	.85
0-7	479 (85.9)	—	—	—
8-10	52 (9.3)	—	—	—
11-21	27 (4.8)	—	—	—
eVAS D (n=552)	—	10.2 (17.3)	0-100	—
0-20	457 (82.8)	—	—	—
21-40	52 (9.4)	—	—	—
41-60	26 (4.7)	—	—	—
61-80	14 (2.5)	—	—	—
81-100	3 (0.6)	—	—	—

[a] Not Applicable.

[b] HADS: Hospital Anxiety and Depression Scale.

[c] eVAS: electronic Visual Analogue Scale.

Table 3. Results for different cutoff scores for anxiety- and depression-screening indices, with STAI-S[a] >39 as a reference for anxiety and MADRS-S[b] >12 as a reference for depression.

Measures	Cutoff score (>)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
HADS[c] anxiety (0-21)					
	4	97	60	59	97
	5	95	71	67	94
	6	88	81	75	90
	7	77	89	87	86
eVAS[d] anxiety (0-100 hs[e])					
	4	91	71	61	94
	5	89	76	63	93
	6	87	78	68	93
	7	86	83	70	93
	8	85	85	72	92
	9	82	86	73	91
	10	79	87	75	90
HADS depression (0-21)					
	4	97	72	56	97
	5	91	83	66	97
	6	87	89	76	94
	7	73	94	81	91
eVAS depression (0-100 hs)					
	4	90	71	48	96
	5	88	77	51	96
	6	85	80	54	96
	7	85	83	56	95
	8	83	84	58	95
	9	82	85	59	95
	10	80	86	60	94

[a] STAI-S: State Trait Anxiety Inventory-State.

[b] MADRS-S: Montgomery Åsberg Depression Rating Scale-Self-report.

[c] HADS: Hospital Anxiety and Depression scale.

[d] eVAS: electronic Visual Analogue Scale.

[e] hs: hundredths.

Discussion

Principal Findings

Web-based versions of eVAS and HADS were excellent with regard to the ability to discriminate between persons with and without clinical levels of anxiety or depression symptoms compared with Web-based versions of MARS-S and STAI-SS. Cutoffs of 5 or 6 on the HADS scales and 7 to 8 hs on the eVAS scales identified patients with anxiety or depression symptoms with high accuracy. Thus, Web-based screening with ultrashort and/or short questionnaires, instead of with longer more burdensome questionnaires, may be sufficient to identify patients with a need for psychosocial support.

Prevalence of Depression and Anxiety

The prevalence of depression and anxiety in this sample is corroborated by studies in similar populations [1,2,47]. This indicates that Web-based screening of anxiety and depression performs equally compared with screening with paper-based questionnaires and may thereby be an adequate screening method in a clinical setting. However, treatment-related consequences may influence the prevalence. A recent review showed that the prevalence of depression among women with breast cancer who had received chemotherapy was higher than among the patients who did not. The risk for depression was higher during the first year after diagnosis, and those receiving adjuvant chemotherapy had higher levels of depression than those who did not. In addition, adverse symptoms of the treatment were associated with decreased health-related quality of life and increased levels of depression [48]. In this study, a majority were under treatment and newly diagnosed, which may have increased anxiety and depression symptoms even though the levels are relatively low.

Determination of Optimal Cutoff Scores of Anxiety and Depression Indices

A questionnaire's ability to detect as many patients with symptoms as possible and to miss as few as possible is of major importance in a first-step screening process. This ability usually comes with the tradeoff of a higher proportion of false positives that must be further evaluated in a more in-depth assessment, which in turn adds extra costs to an already strained health care sector [49]. Thus, there is a delicate balance between an excessively high and excessively low cutoff score from a clinical point of view.

Anxiety

For HADS anxiety, a cutoff of >5 may be suggested as it has the highest ability to correctly identify a patient with clinical symptoms of anxiety, weighing in both sensitivity and specificity. Previous Swedish studies in similar settings found that a cutoff of >4 on HADS anxiety was the best to detect the patients who deteriorated in their psychological health and quality of life during the disease trajectory [32,50]. However, as declared in the method section, there is still no consensus regarding the optimal cutoff for any of the HADS subscales for the paper-based format, indicating that future evaluations of the Web-based version of HADS will also yield various recommendations regarding cutoff scores. A cutoff score of >7 hs on eVAS anxiety may be recommended as this score has a high sensitivity, according to our results. In a previous comparison study of the eVAS anxiety and STAI-S, it was found that the 1-item eVAS could be readily completed and that it adequately measured anxiety [51]. Yet, further evaluations of the eVAS are needed before firm conclusions regarding an optimal cutoff can be drawn.

Depression

As an initial step in a screening procedure, a cutoff score of >5 on HADS depression may be preferable as it has a high ability to identify a patient with clinical symptoms of depression as well as to avoid failure detecting individuals without symptoms of depression. The HADS is the most thoroughly evaluated measurement of depression in oncology settings but there are difficulties comparing findings with earlier studies as many cutoffs have been used. Even lower cutoffs have been recommended for patients with cancer [31,52], and consensus about the optimal cutoff is still not reached. Regarding the eVAS, a cutoff score of >7 hs may be most suitable as it has a high ability to identify a patient with depression symptoms and rule out the ones without, but again, this need to be further evaluated.

Strength and Limitations

This study has some limitations. First, the sample was restricted to newly diagnosed patients with breast, prostate, or colorectal cancer or a colorectal cancer relapse. Thus, the results may not represent patients with other cancer diagnoses or long-term survivors. On the contrary, we have studied a group of both curable and noncurable patients, which increases the probability that the results could be generalized to a bigger group than the one in this study. Another limitation is that the data are collected within a longitudinal intervention study that is much more demanding than a study where the participants are asked to answer questionnaires at a single observation point. This may have affected the sample and the number of nonparticipants and withdrawals (n=939). Our results need to be confirmed by future studies designed to compare Web-based instruments. In addition, we did not include patients who were not internet users and participants who declined Web-based studies, commonly representing an important group of older, sicker, and less-educated patients [53].

There may be a problem with face validity with regard to the Web-based version of the eVAS in this study as the mean score was low compared with the paper and pen versions [23]. However, mean ratings for eVAS have been shown to be equal when comparing the value received from eVAS with the one gathered via other formats in Web-based studies [54]. Another explanation of the low scores on the eVAS could be the wording of the questions. In this study, we used the expressions *anxiety* and *depression*, which may have been interpreted as a relatively severe condition compared with *worry* and *low mood*, which are more commonly used in Swedish. This choice of wording may have led to a higher number of true positives but also to more undetected patients with symptoms. This needs to be further elaborated upon to enable any certain conclusions.

We used the MADRS-S and STAI-S as reference instruments for depression and anxiety, which is not beyond criticism, given that the MADRS-S and STAI-S are not the most validated instruments in cancer settings. However, they have both shown good psychometric properties in various studies in different settings, including cancer and Web-based settings [40,55,56]. The ability to detect patients at suicidal risk is documented in the Web-based version of MADRS-S [40], which is of major importance in Web-based screening. The choice of self-report measures to establish the expected prevalence of anxiety and depression as well as relevant cutoffs should be interpreted with caution, whereas in-depth clinical interviews would have been more reliable.

The internal consistency was good in all 4 instruments indicated by high Cronbach alpha values. The sample size was good given that 558 patients responded to the questionnaires.

Clinical Implications

This study indicates that the ultrashort and short eVAS and HADS may be used in a Web-based screening procedure in cancer care. Thus, a short screening instrument for anxiety and depression may be suggested in the initial step of a screening process, as they are less demanding and time-consuming. However, only 37.3% (558/1497) of eligible patients completed the Web-based questionnaires in this study. One of the main reasons for this was a lack of internet access. This makes it clear that Web-based screening strategies must include easy access to touch screens or tablets and support for the use of such devices. Both the eVAS and HADS perform well with regard to the ability to identify a majority of patients with anxiety and depression symptoms. The problem with false positives can be solved by further Web-based assessments of the patients reporting symptoms to allocate the support adequately. Some studies recommend the HADS as a follow-up instrument to detect cases after initial screenings with an ultrashort instrument [57], whereas others demonstrate that the HADS is more suitable as an instrument for initial screenings of depression and anxiety and that it cannot be recommended as a diagnostic (case-finding) instrument [17,58]. Patients reporting symptoms in the initial screening process may be further assessed with instruments with a higher specificity and must always be followed by a clinical assessment for diagnostics and treatment decisions.

Suggestions for Future Research

The existing Web-based instruments are often paper questionnaires that have been adapted for Web-based use. The correlation between paper and Web-based versions of the same instruments correlate strongly, but differences such as mean scores and psychometrics do appear [22]. Further research on the Web-based versions is crucial to evaluate the reliability and validity of the existing instruments. Development of a new Web-based screening instrument may be optimal and should be considered in future research. This study has contributed to the field by demonstrating that the eVAS and HADS are comparable with longer questionnaires with regard to identification of patients with anxiety and depression symptoms and may be used as an initial Web-based screening instrument among patients with cancer. However, screening in itself is not sufficient. It has to take place in an infrastructure comprising adequate support and follow-up, and studies regarding the prerequisite for successful implementation in a clinical setting are needed.

Conclusions

The use of the short and ultrashort tools eVAS and HADS is a suitable initial method of Web-based screening in busy clinical settings. However, there are still a large proportion of patients who lack access to the internet or the ability to use it. There is a need for special solutions for this group to find all the patients with psychological distress.

Conflicts of Interest

None declared.

References

1. Linden W, Vodermaier A, Mackenzie R, Greig D. Anxiety and depression after cancer diagnosis: prevalence rates by cancer type, gender, and age. J Affect Disord 2012 Dec 10;141(2-3):343-351. [doi: 10.1016/j.jad.2012.03.025] [Medline: 22727334]
2. Mitchell AJ. Screening for cancer-related distress: when is implementation successful and when is it unsuccessful? Acta Oncol 2013 Feb;52(2):216-224. [doi: 10.3109/0284186X.2012.745949] [Medline: 23320770]
3. Thalén-Lindström A, Larsson G, Glimelius B, Johansson B. Anxiety and depression in oncology patients; a longitudinal study of a screening, assessment and psychosocial support intervention. Acta Oncol 2013 Jan;52(1):118-127. [doi: 10.3109/0284186X.2012.707785] [Medline: 22934559]
4. Caruso R, Nanni MG, Riba M, Sabato S, Mitchell AJ, Croce E, et al. Depressive spectrum disorders in cancer: prevalence, risk factors and screening for depression: a critical review. Acta Oncol 2017 Feb;56(2):146-155. [doi: 10.1080/0284186X.2016.1266090] [Medline: 28140731]
5. Smith HR. Depression in cancer patients: pathogenesis, implications and treatment (review). Oncol Lett 2015 Apr;9(4):1509-1514 [FREE Full text] [doi: 10.3892/ol.2015.2944] [Medline: 25788991]
6. Satin JR, Linden W, Phillips MJ. Depression as a predictor of disease progression and mortality in cancer patients. Cancer 2009 Sep 14;115(22):5349-5361. [doi: 10.1002/cncr.24561]
7. Arrieta O, Angulo LP, Núñez-Valencia C, Dorantes-Gallareta Y, Macedo EO, Martínez-López D, et al. Association of depression and anxiety on quality of life, treatment adherence, and prognosis in patients with advanced non-small cell lung cancer. Ann Surg Oncol 2013 Jun;20(6):1941-1948. [doi: 10.1245/s10434-012-2793-5] [Medline: 23263699]
8. Andersen BL, Rowland JH, Somerfield MR. Screening, assessment, and care of anxiety and depressive symptoms in adults with cancer: an american society of clinical oncology guideline adaptation. J Oncol Pract 2015 Mar;11(2):133-134. [doi: 10.1200/JOP.2014.002311] [Medline: 25515721]
9. Mitchell AJ, Kaar S, Coggan C, Herdman J. Acceptability of common screening methods used to detect distress and related mood disorders-preferences of cancer specialists and non-specialists. Psychooncology 2008 Mar;17(3):226-236. [doi: 10.1002/pon.1228] [Medline: 17575565]
10. Faller H, Weis J, Koch U, Brähler E, Härter M, Keller M, et al. Utilization of professional psychological care in a large German sample of cancer patients. Psychooncology 2017 Dec;26(4):537-543. [doi: 10.1002/pon.4197] [Medline: 27327213]
11. National Institute for Clinical Excellence. London; 2004. Improving supportive and palliative care for adults with cancer URL: https://www.nice.org.uk/guidance/csg4 [accessed 2019-02-25] [WebCite Cache ID 76RoYqAa1]
12. Holland JC, Bultz BD, National comprehensive Cancer Network (NCCN). The NCCN guideline for distress management: a case for making distress the sixth vital sign. J Natl Compr Canc Netw 2007 Jan;5(1):3-7. [Medline: 17323529]

13. Yu ES, Shim EJ, Kim HK, Hahm BJ, Park JH, Kim JH. Development of guidelines for distress management in Korean cancer patients. Psycho-Oncology 2012;21(5):541-549 [FREE Full text] [doi: 10.1002/pon.1931] [Medline: 21351189]
14. Salmon P, Clark L, McGrath E, Fisher P. Screening for psychological distress in cancer: renewing the research agenda. Psychooncology 2015 Mar;24(3):262-268. [doi: 10.1002/pon.3640] [Medline: 25082459]
15. Goodyear-Smith FA, van Driel ML, Arroll B, Del Mar C. Analysis of decisions made in meta-analyses of depression screening and the risk of confirmation bias: a case study. BMC Med Res Methodol 2012 Jun 12;12:76 [FREE Full text] [doi: 10.1186/1471-2288-12-76] [Medline: 22691262]
16. Bidstrup PE, Johansen C, Mitchell AJ. Screening for cancer-related distress: summary of evidence from tools to programmes. Acta Oncol 2011 Feb;50(2):194-204. [doi: 10.3109/0284186X.2010.533192] [Medline: 21231781]
17. Mitchell AJ. Short screening tools for cancer-related distress: a review and diagnostic validity meta-analysis. J Natl Compr Canc Netw 2010 Apr;8(4):487-494. [Medline: 20410338]
18. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. J Natl Cancer Inst 2009 Nov 04;101(21):1464-1488 [FREE Full text] [doi: 10.1093/jnci/djp336] [Medline: 19826136]
19. Leykin Y, Thekdi SM, Shumay DM, Muñoz RF, Riba M, Dunn LB. Internet interventions for improving psychological well-being in psycho-oncology: review and recommendations. Psychooncology 2012 Sep;21(9):1016-1025 [FREE Full text] [doi: 10.1002/pon.1993] [Medline: 21608075]
20. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
21. Carlson LE, Groff SL, Maciejewski O, Bultz BD. Screening for distress in lung and breast cancer outpatients: a randomized controlled trial. J Clin Oncol 2010 Nov 20;28(33):4884-4891. [doi: 10.1200/JCO.2009.27.3698] [Medline: 20940193]
22. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014 Dec 03;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
23. Cella DF, Perry SW. Reliability and concurrent validity of three visual-analogue mood scales. Psychol Rep 1986 Oct;59(2 Pt 2):827-833. [doi: 10.2466/pr0.1986.59.2.827] [Medline: 3809343]
24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
25. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [Medline: 444788]
26. Spielberger C, Gorsuch R, Lushene R. Manual for the State-Trait Anxiety (STAI Form Y). Palo Alto, CA: Consulting Psychologist Press; 1983.
27. Mattsson S, Alfonsson S, Carlsson M, Nygren P, Olsson E, Johansson B. U-CARE: internet-based stepped care with interactive support and cognitive behavioral therapy for reduction of anxiety and depressive symptoms in cancer--a clinical trial protocol. BMC Cancer 2013 Sep 11;13:414 [FREE Full text] [doi: 10.1186/1471-2407-13-414] [Medline: 24024826]
28. Jacobsen PB, Donovan KA, Trask PC, Fleishman SB, Zabora J, Baker F, et al. Screening for psychologic distress in ambulatory cancer patients. Cancer 2005 Apr 01;103(7):1494-1502 [FREE Full text] [doi: 10.1002/cncr.20940] [Medline: 15726544]
29. Carlson LE, Waller A, Groff SL, Zhong L, Bultz BD. Online screening for distress, the 6th vital sign, in newly diagnosed oncology outpatients: randomised controlled trial of computerised vs personalised triage. Br J Cancer 2012 Aug 07;107(4):617-625 [FREE Full text] [doi: 10.1038/bjc.2012.309] [Medline: 22828610]
30. Sellick SM, Edwardson AD. Screening new cancer patients for psychological distress using the hospital anxiety and depression scale. Psychooncology 2007 Jun;16(6):534-542. [doi: 10.1002/pon.1085] [Medline: 16989004]
31. Vodermaier A, Millman RD. Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. Support Care Cancer 2011 Dec;19(12):1899-1908. [doi: 10.1007/s00520-011-1251-4] [Medline: 21898134]
32. Thalén-Lindström AM, Glimelius BG, Johansson BB. Identification of distress in oncology patients: a comparison of the Hospital Anxiety and Depression scale and a thorough clinical assessment. Cancer Nurs 2016;39(2):E31-E39. [doi: 10.1097/NCC.0000000000000267] [Medline: 26018818]
33. Carey M, Noble N, Sanson-Fisher R, MacKenzie L. Identifying psychological morbidity among people with cancer using the Hospital Anxiety and Depression Scale: time to revisit first principles? Psychooncology 2012 Mar;21(3):229-238. [doi: 10.1002/pon.2057] [Medline: 21919120]
34. Verdonck-de LI, de Bree R, Keizer A, Houffelaar T, Cuijpers P, van der Linden MH. Computerized prospective screening for high levels of emotional distress in head and neck cancer patients and referral rate to psychosocial care. Oral oncology Oct 2009;45(10):e129-e133. [doi: 10.1016/j.oraloncology.2009.01.012] [Medline: 19362038]
35. Cronly J, Duff AJ, Riekert KA, Perry IJ, Fitzgerald AP, Horgan A. Online versus paper-based screening for depression and anxiety in adults with cystic fibrosis in Ireland: a cross-sectional exploratory study. BMJ open 2018;8(1):e019305. [doi: 10.1136/bmjopen-2017-019305] [Medline: 29358445]

36. Brown Johnson CG, Brodsky J, Cataldo J. Lung cancer stigma, anxiety, depression, and quality of life. J Psychosoc Oncol 2014;32(1):59-73 [FREE Full text] [doi: 10.1080/07347332.2013.855963] [Medline: 24428251]
37. Peoples AR, Bushunow PW, Garland SN, Heckler CE, Roscoe JA, Peppone LL, et al. Buspirone for management of dyspnea in cancer patients receiving chemotherapy: a randomized placebo-controlled URCC CCOP study. Support Care Cancer 2016 Mar;24(3):1339-1347 [FREE Full text] [doi: 10.1007/s00520-015-2903-6] [Medline: 26329396]
38. Knight RG, Waal-Manning HJ, Spears GF. Some norms and reliability data for the State--Trait Anxiety Inventory and the Zung Self-Rating Depression scale. Br J Clin Psychol 1983 Nov;22 (Pt 4):245-249. [Medline: 6640176]
39. Fantino B, Moore N. The self-reported Montgomery-Asberg Depression Rating scale is a useful evaluative tool in major depressive disorder. BMC Psychiatry 2009 May 27;9:26 [FREE Full text] [doi: 10.1186/1471-244X-9-26] [Medline: 19473506]
40. 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]
41. Svanborg P, Ekselius L. Self-assessment of DSM-IV criteria for major depression in psychiatric out- and inpatients. Nord J Psychiatry 2003;57(4):291-296. [Medline: 12888397]
42. Hellsten C, Sjöström K, Lindqvist PG. A prospective Swedish cohort study on psychosocial factors influencing anxiety in women referred for colposcopy. BJOG 2007 Jan;114(1):32-38 [FREE Full text] [doi: 10.1111/j.1471-0528.2006.01161.x] [Medline: 17233857]
43. Wikberg C, Nejati S, Larsson ME, Petersson E, Westman J, Ariai N, et al. Comparison between the Montgomery-Asberg Depression Rating Scale-Self and the Beck Depression Inventory II in primary care. Prim Care Companion CNS Disord 2015;17(3):- [FREE Full text] [doi: 10.4088/PCC.14m01758] [Medline: 26644958]
44. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
45. IBM Corp. New York; 2011. IBM SPSS statistics for Windows, version 20 URL: https://www.ibm.com/products/spss-statistics [accessed 2019-02-27] [WebCite Cache ID 76Um6yBtF]
46. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4(2):627-635 [FREE Full text] [Medline: 24009950]
47. Herschbach P, Berg P, Dankert A, Duran G, Engst-Hastreiter U, Waadt S, et al. Fear of progression in chronic diseases: psychometric properties of the Fear of Progression Questionnaire. J Psychosom Res 2005 Jun;58(6):505-511. [doi: 10.1016/j.jpsychores.2005.02.007] [Medline: 16125517]
48. Fann JR, Thomas-Rich AM, Katon WJ, Cowley D, Pepping M, McGregor BA, et al. Major depression after breast cancer: a review of epidemiology and treatment. Gen Hosp Psychiatry 2008;30(2):112-126. [doi: 10.1016/j.genhosppsych.2007.10.008] [Medline: 18291293]
49. Mitchell AJ. Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders. J Clin Oncol 2007 Oct 10;25(29):4670-4681. [doi: 10.1200/JCO.2006.10.0438] [Medline: 17846453]
50. Thalén-Lindström A, Glimelius B, Johansson B. Development of anxiety, depression and health-related quality of life in oncology patients without initial symptoms according to the Hospital Anxiety and Depression Scale - a comparative study. Acta Oncol 2017 Aug;56(8):1094-1102. [doi: 10.1080/0284186X.2017.1305124] [Medline: 28339305]
51. Davey HM, Barratt AL, Butow PN, Deeks JJ. A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. J Clin Epidemiol 2007 Apr;60(4):356-360. [doi: 10.1016/j.jclinepi.2006.07.015] [Medline: 17346609]
52. Singer S, Kuhnt S, Götze H, Hauss J, Hinz A, Liebmann A, et al. Hospital anxiety and depression scale cutoff scores for cancer patients in acute care. Br J Cancer 2009 Mar 24;100(6):908-912 [FREE Full text] [doi: 10.1038/sj.bjc.6604952] [Medline: 19240713]
53. Girault A, Ferrua M, Lalloué B, Sicotte C, Fourcade A, Yatim F, et al. Internet-based technologies to improve cancer care coordination: current use and attitudes among cancer patients. Eur J Cancer 2015 Mar;51(4):551-557. [doi: 10.1016/j.ejca.2014.12.001] [Medline: 25661828]
54. Kuhlmann T, Reips U, Wienert J, Lippke S. Using Visual Analogue scales in eHealth: non-response effects in a lifestyle intervention. J Med Internet Res 2016 Dec 22;18(6):e126 [FREE Full text] [doi: 10.2196/jmir.5271] [Medline: 27334562]
55. Ketelaars PJ, Buskes MH, Bosgraaf RP, van Hamont D, Prins JB, Massuger LF, et al. The effect of video information on anxiety levels in women attending colposcopy: a randomized controlled trial. Acta Oncol 2017 Dec;56(12):1728-1733. [doi: 10.1080/0284186X.2017.1355108] [Medline: 28760058]
56. Rhondali W, Freyer G, Adam V, Filbet M, Derzelle M, Abgrall-Barbry G, et al. Agreement for depression diagnosis between DSM-IV-TR criteria, three validated scales, oncologist assessment, and psychiatric clinical interview in elderly patients with advanced ovarian cancer. Clin Interv Aging 2015;10:1155-1162 [FREE Full text] [doi: 10.2147/CIA.S71690] [Medline: 26203235]

57. Butow P, Price MA, Shaw JM, Turner J, Clayton JM, Grimison P, et al. Clinical pathway for the screening, assessment and management of anxiety and depression in adult cancer patients: Australian guidelines. Psychooncology 2015 Sep;24(9):987-1001. [doi: 10.1002/pon.3920] [Medline: 26268799]
58. Luckett T, Butow PN, King MT, Oguchi M, Heading G, Hackl NA, et al. A review and recommendations for optimal outcome measures of anxiety, depression and general distress in studies evaluating psychosocial interventions for English-speaking adults with heterogeneous cancer diagnoses. Support Care Cancer 2010 Oct;18(10):1241-1262. [doi: 10.1007/s00520-010-0932-8] [Medline: 20596731]

Abbreviations

AUC: area under the curve
eVAS: electronic Visual Analogue Scale
HADS: Hospital Anxiety and Depression Scale
MADRS-S: Montgomery Åsberg Depression Rating Scale-Self-report
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating curve
STAI-S: State Trait Anxiety Inventory-State
VAS: Visual Analogue Scale

Edited by G Eysenbach; submitted 25.06.18; peer-reviewed by K Donovan, T Clinton-McHarg; comments to author 07.10.18; revised version received 18.01.19; accepted 07.02.19; published 05.04.19.

Please cite as:
Mattsson S, Olsson EMG, Carlsson M, Johansson BBK
Identification of Anxiety and Depression Symptoms in Patients With Cancer: Comparison Between Short and Long Web-Based Questionnaires
J Med Internet Res 2019;21(4):e11387
URL: https://www.jmir.org/2019/4/e11387/
doi:10.2196/11387
PMID:30950804

Original Paper

Sociodemographic, Attitudinal, and Behavioral Correlates of Using Nutrition, Weight Loss, and Fitness Websites: An Online Survey

Carlos A Almenara[1*], BSc, MSc, PhD; Hana Machackova[2,3*], MA, PhD; David Smahel[2,4*], PhD

[1]Faculty of Psychology, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
[2]Institute for Research on Children, Youth and Family (IVMDR), Department of Psychology, Masaryk University, Brno, Czech Republic
[3]Department of Media Studies and Journalism, Masaryk University, Brno, Czech Republic
[4]Faculty of Informatics, Masaryk University, Brno, Czech Republic
[*]all authors contributed equally

Corresponding Author:
Carlos A Almenara, BSc, MSc, PhD
Faculty of Psychology
Universidad Peruana de Ciencias Aplicadas
Avenida Alameda San Marcos, s/n
Lima, 15067
Peru
Phone: 51 3133333 ext 2803
Email: carlos.almenara@upc.pe

Abstract

Background: Nutrition, diet, and fitness are among the most searched health topics by internet users. Besides that, health-related internet users are diverse in their motivations and individual characteristics. However, little is known about the individual characteristics associated with the usage of nutrition, weight loss, and fitness websites.

Objective: The aim of this study was to examine the individual factors associated with the usage of nutrition, weight loss, and fitness websites.

Methods: An invitation to an online survey was published on 65 websites and discussion forums. In total, we employed data from 623 participants (aged 13 to 39 years, mean 24.11 [SD 5.26]). The measures included frequency of usage of nutrition, weight loss and fitness websites, excessive exercise, eating disorder symptomatology, internalization of the beauty ideal, weight status, and perceived online social support. Participants' data were used as predictors in a base linear regression model.

Results: The final model had an acceptable fit (χ^2_{10} =14.1; P=.17; root mean square error of approximation=0.03; comparative fit index=0.99; Tucker-Lewis index=0.99). Positive associations were found between usage of (1) nutrition websites and being female, higher levels of excessive exercise, and perceived online social support; (2) weight loss websites and excessive exercise, internalization, being female, eating disorder symptomatology, and being overweight or obese; and (3) fitness websites and levels of excessive exercise, internalization, and frequency of internet use.

Conclusions: The results highlighted the importance of individual differences in the usage of health-related websites.

(J Med Internet Res 2019;21(4):e10189) doi:10.2196/10189

KEYWORDS

body image; compulsive behavior; diet; feeding and eating disorders; individuality; internet; user-computer interface; weight loss; social support

Introduction

Background

Nutrition, diet, and fitness are among the most searched health topics by internet users [1]. The usage of websites covering those topics can be related to positive outcomes, such as a healthier diet or higher adherence to physical activity [2]. However, it can also be associated with negative outcomes such as unhealthy weight control behaviors [3,4]. This is worrisome considering that some internet users could accentuate their eating and weight-related problems (eg, eating disorders and obesity).

Notable efforts have been made to understand health-related internet users [2,5,6]. However, no study has empirically tested a group of already identified individual characteristics related with eating and weight-related problems and its link with the usage frequency of nutrition, weight loss, and fitness websites. This study is aimed to fill this gap because of the need for a better understanding of the psychological factors related to health behaviors, something crucial to develop successful interventions [7]. Moreover, by understanding which individual factors are associated with each of these websites, we can improve the design and communication of tailored health information. In other words, we will be able to provide more personalized information within nutrition, weight loss, and fitness websites to foster health-enhancing behaviors rather than health-compromising behaviors.

Past Research

Health-related internet users are diverse in their motivations and individual characteristics [8]. From past research, we have identified a group of individual characteristics associated with the selection and usage frequency of nutrition, weight loss, and fitness websites in different studies: eating disorder symptomatology [4]; the extent of the internalization of the ideal body [9]; the levels of exercise [3]; social support [10]; body mass index (BMI); and sociodemographic characteristics, such as gender and age [2,11].

For instance, studies on internet usage regarding diet, weight loss, and fitness have found that as many as 85% of the users are female [12-14]. Notably, studies with young women have found that the use of Web-based weight loss information predicts disordered eating [4], whereas self-reported eating disorder also predicts the interest in diet and fitness websites [13]. Therefore, it is common to see individuals with disordered eating symptomatology joining online communities mostly to get tips and tricks for weight loss, as well as social support [15]. Similarly, overweight and particularly obese individuals who experience the stigma against obesity [16] commonly turn to the internet for weight loss solutions [17].

Internalization is another important individual factor. There is evidence that media spreads an ideal of beauty that people *internalize* and renders them more prone to inappropriately change their nutrition or physical activity habits, such as doing strenuous exercise to achieve the “ideal body” [18,19]. In this sense, it is suggested that individuals with high levels of internalization (ie, susceptible individuals) would gravitate to appearance-focused media content, such as websites focused on weight loss for appearance reasons. Following Perloff’s model, it should be noted that, rather than a one-way effect of media use on attitudes and behaviors, there would be a mutual reinforcement [9]. In other words, internalization not only accounts for the effect of media exposure [19] but also for a dispositional factor to subsequent media usage [20].

Although there is some evidence for the influence of online media on exercise and fitness [3], less is known about the association between excessive exercise and the usage of healthy lifestyle websites. For instance, a recent study exploring the characteristics of women who post travel images and women who post “fitspiration” images (ie, promoting a healthy lifestyle through fitness) found higher levels of compulsive exercise and disordered eating in the fitspiration group [21]. In contrast, another study found that the usage of mobile phone apps to keep track of meals or exercise routines, but not blogs or microblogging (eg, Instagram and Twitter) about nutrition and exercise, was associated with compulsive exercise [22].

Finally, large studies suggest that internet usage for health information is highly motivated by the opportunity to get advice from others [6]. For instance, a study of an online community of individuals with high levels of disordered eating found that the main motive to join the group was to get support and advice regarding weight loss [15]. Similarly, other studies suggest that some people turn to the internet to supplement professional medical advice, particularly when they are looking for advice on specific health issues or conditions [10,17]. Thus, online peer-to-peer health care reflects the importance of the online social support for adults’ health behaviors [7,10].

This Study

Consequently, this study was aimed at exploring which individual factors relevant for the prevention of eating and weight-related problems are associated with the frequency of the usage of nutrition, weight loss, and fitness websites. We specifically examined the links with sociodemographic variables (ie, gender and age), eating disorder symptomatology, weight status, the tendency for excessive exercise, and the levels of internalization and perceived social support from website users. Upon this examination, we aimed to uncover which of the abovementioned factors could predict the usage of those websites. The results can be helpful in the design of tailored health messages for the prevention of eating and weight-related problems and the promotion of a healthy lifestyle.

Methods

Participants

The study utilized data from the visitors of websites focused on nutrition, weight loss, and exercise collected as part of a project on eating behaviors in the context of internet and technology use. It was approved by the Research Ethics Committee of Masaryk University. The data were collected through an open Web survey between May and October 2016. For participant recruitment, Czech websites oriented to nutrition, weight loss, and fitness were asked to publish an invitation for study participation. These included websites, Web magazines, blogs, social networking sites, and specific discussions forums which were all searched using keywords related to nutrition, weight loss, and fitness. In total, 307 different online platforms were asked to publish the invitation to the survey. The invitation was published on 65 websites and discussion forums (with a response rate of 21%). All the participants were informed about the purpose of the research, the estimated time needed to fill in the questionnaire, and the right to the questionnaire at any time were provided link to further information about the project; and were asked to provide informed consent by clicking on the link to the questionnaire. Participants were motivated by the chance to win 1 of 5 vouchers for an e-shop for the amount of 40 Euros each. From the original sample, which comprises 1002 respondents (age mean 24.82 (SD 6.85); 81.64% [818/1002] females), we excluded (1) participants aged 40 years and older because of low number of respondents in this age (3.59%, 36/1002) and to keep our sample more homogeneous in terms of age; (2) participants who did not provide a sufficient amount of data regarding their individual characteristics which were measured on the last page on the questionnaire (27.34%, 274/1002); and (3) participants who reported that the reason for the website visits was because of the health issues of someone else (therefore, possibly lacking the personal motivation connected to their own eating and health status) (5.49%, 55/1002). The latter reason for exclusion was indicated by the question “Do you visit the sites about nutrition or sports not for yourself, but mainly because you want to help with the nutrition or sport of another person (partner, child, parent, etc.)?” and the answer “Definitely applies”. This excluded 335 respondents in total. Moreover, we excluded respondents with occasional missing values (4.39%, 44/1002), yielding a final sample of 623 respondents aged 13 to 39 years (mean 24.11, SD 5.26), including 83.6% females (521/623). The majority of our respondents were Czech (91.7%, 571/623); 7.7% were Slovak (48/623); and 0.6% indicated other nationality (4/623). All of these respondents were internet users who go online “several times a week” (0.9%, 6/623), “almost daily” (9.9%, 62/623), and “daily” (88.9%, 554/623).

Measures

Usage of Nutrition, Weight Loss, and Fitness Websites

The frequency of use was measured by the question “How often do you visit websites regarding nutrition, weight loss, or exercise and sport?” with answers on a 6-point scale with the response options as follows: 1 (Never), 2 (Almost never), 3 (Several times a month), 4 (Several times a week), 5 (Almost daily), and 6 (Daily). Respondents answered with regard to 3 types of websites, that is, those focused on *nutrition (eg, relating to specific diets and healthy meals)* (mean 4.38, SD 1.23); *weight loss (eg, diets or instructions on how to lose weight)* (mean 3.04, SD 1.45); and *fitness (regarding your exercise or sport, but not, eg, the results of professional athletes)* (mean 4.02, SD 1.39).

Gender and Age

Gender was coded in binary (0=males, 1=females), and age was also requested (mean 24.11, SD 5.26 years).

Excessive Exercise

A total of 5 items from the excessive exercise subscale from the Eating Pathology Symptoms Inventory Scales [23] answered on a 5-point scale ranging from 1 (Never) to 5 (Very often) were used. The scale was computed by averaging the items, with higher scores indicating greater tendency for excessive exercise (mean 3.05, SD 0.96; alpha=.87); factor analysis confirmed unidimensional structure, all loadings >.75.

Eating Disorder Symptomatology

The SCOFF screening tool [24] consisting of 5 items with Yes/No response options was used to identify a group of respondents potentially at risk of eating disorders. Those answering “Yes” on 2 or more items were classified as respondents at risk of having an eating disorder (47.2%, 294/623).

Weight Status

The respondents reported their current height (in centimeters) and weight (in kilograms), which were used to calculate their BMI (kg/m^2) (mean 23.02, SD 4.24). Weight status data were then obtained using international cut-off points for adults [25] and adolescents [26]. Respondents were classified as either being *underweight* (6.4%, 40/623), having *normal weight* (70.3%, 438/623), and being *overweight* or *obese* (23.3%, 145/623).

Internalization of the Beauty Ideal

Respondents were asked *to what extent do the following statements apply to you in regards to these sites?*, with 3 items adapted from the Multidimensional Media Influence Scale [27]: *I am comparing my appearance with people on these sites*, *I am trying to look like the people on these sites*, and *the content on these sites inspire me in how to look attractive*. The items were answered on a 4-point scale ranging from 1 (*definitely does not apply*) to 4 (*definitely applies*). The scale was computed by averaging the items, with higher scores indicating higher internalization (mean 2.34, SD 0.90; alpha=.82); factor analysis confirmed unidimensional structure, all loadings >.83.

Perceived Online Social Support

Respondents were asked “To what extent do the following statements apply to you with regard to these sites?”, with 3 items adapted from the Online Social Support for Smokers Scale [28]: “I get advice and support here that I would not get elsewhere”, “It is encouraging to know that there are other people making similar efforts (with regard to nutrition or sport)”, and “I feel that other visitors (or authors) of sites are giving me support”. The items were answered on a 4-point scale ranging from 1 (Definitely does not apply) to 4 (Definitely applies). The scale was computed by averaging the items, with higher scores indicating higher perceived support (mean 2.80, SD 0.74; alpha=.72); factor analysis confirmed unidimensional structure, all loadings >.70.

Statistical Analysis

To assess the links between the studied factors, we tested a base linear regression model with observed variables in which we predicted all 3 outcomes using MLR estimator in Mplus (version 7; Muthén & Muthén, Los Angeles, CA, USA). All paths between predictors and outcomes were allowed; weight status was included as a dummy variable with *normal weight* as reference category. In the second step, we constrained all nonsignificant (*P*<.05) paths to zero.

Results

Results of the base linear regression model are shown below (Table 1).

The final model (Figure 1) had an acceptable fit (χ^2_{10}=14.0, *P*=.17, root mean square error of approximation=0.03, comparative fit index=0.99, Tucker-Lewis index=0.99). The more frequent visits of nutrition websites were positively predicted by being female and having higher excessive exercise and higher perceived online social support. The visits of weight loss websites were predicted by being female and having higher excessive exercise, higher internalization, higher perceived online social support, eating disorder symptomatology, and being overweight (as opposed to having normal weight); the effect of age was significant but negligible (*beta*=.082). The visits of fitness-oriented websites were positively predicted only by higher excessive exercise and higher internalization; the effect of age was again significant but negligible (*beta*=.095), and the effect of perceived online social support was not significant in the final model (*P*=.078). Owing to the low number of respondents in the "underweight" category, we ran the analysis as well without this category; nevertheless, this supplemental analysis did not yield any substantially different results.

Table 1. Base regression model predicting the frequency of the usage of nutrition, weight loss, and fitness websites.

Variables	Nutrition[a]		Weight loss[b]		Fitness[c]	
	Beta	*P* value	Beta	*P* value	Beta	*P* value
Gender (females)	0.211	<.001	0.184	<.001	0.004	0.9
Age (years)	0.003	0.93	0.082	0.01	0.092	0.006
Eating disorder symptomatology	0.053	0.15	0.165	<.001	−.049	0.14
Excessive exercise	0.277	<.001	0.257	<.001	0.57	<.001
BMI[d]—underweight (vs normal)	0.066	0.06	−.003	0.94	0.032	0.35
BMI—overweight (vs normal)	0.026	0.49	0.152	<.001	0.011	0.73
Internalization	−.079	0.07	0.19	<.001	0.129	0.001
Perceived online social support	0.32	<.001	0.11	0.002	0.063	0.049

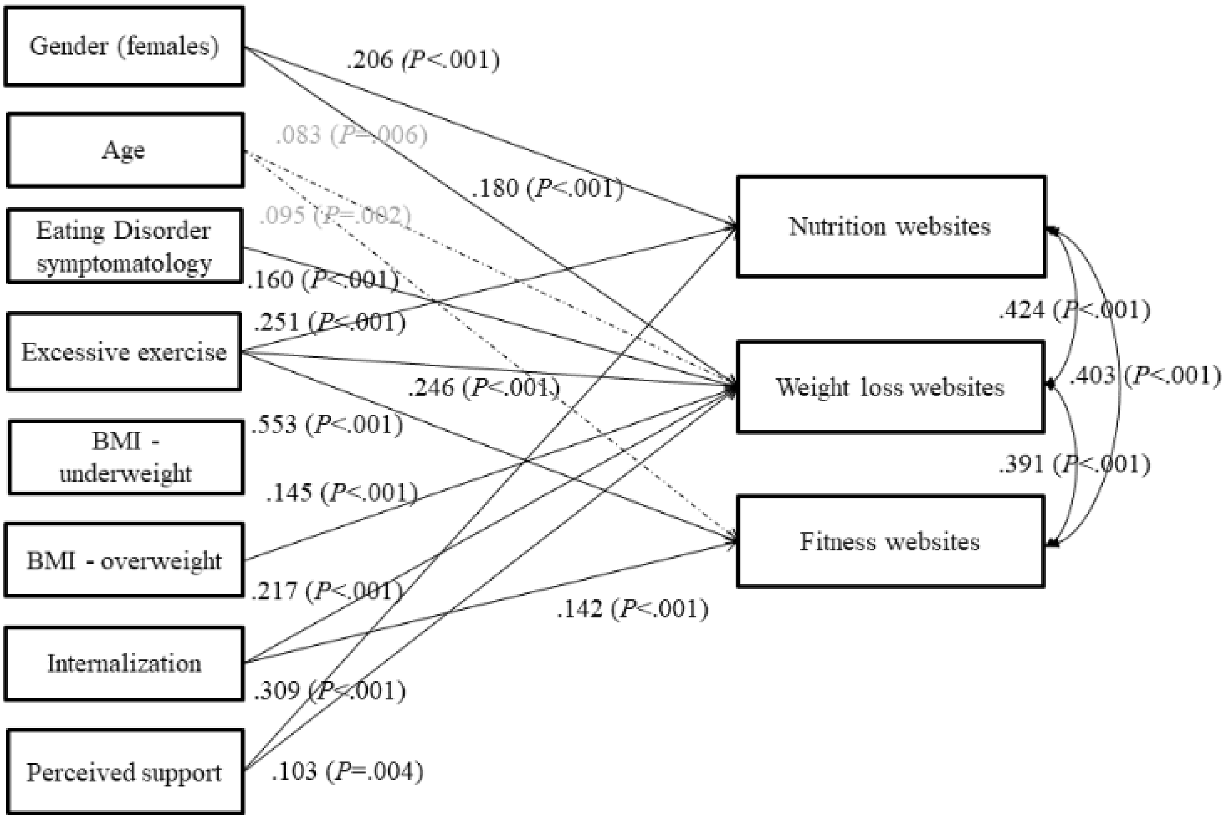
[a] Mean (SD) 4.38 (1.23); R^2=.223.

[b] Mean (SD) 3.04(1.45); R^2=.286.

[c] Mean (SD) 4.02(1.39); R^2=.400.

[d] BMI: body mass index.

Figure 1. Final model predicting the frequency of the usage of nutrition, weight loss, and fitness websites. Standardized coefficients; effects with beta<.01 are dashed; the effect of perceived support on visits of fitness websites is not shown as it is not significant (beta=.056; *P*=.078). BMI: body mass index.



Discussion

Principal Findings

The aim of this study was to explore which individual factors relevant for the prevention of eating and weight-related problems were associated with the frequency of the usage of nutrition, weight loss, and fitness websites. Compared with previous studies that have examined separately those individual characteristics regarding the usage of those websites separately, this study was the first one to examine all those individual factors together regarding the usage of each of those 3 types of websites. Therefore, our study brings the opportunity to compare how an equal group of individual characteristics relate differently to the usage of nutrition, weight loss, and fitness websites. These results have several implications, as we show next.

Being female, having higher levels of excessive exercise, and perceived online social support were positively associated with the usage of nutrition websites. The usage of weight loss websites was positively associated with excessive exercise, internalization, being female, eating disorder symptomatology, and being overweight or obese, whereas its association with age and online social support was very weak. The frequency of fitness website usage was associated with the levels of excessive exercise, internalization (although this association was lower), and age (although this association was close to zero). In general terms, these findings contribute to a better understanding of the online health behavior of users of nutrition, weight loss, and fitness websites. Moreover, these results can contribute with future initiatives for the prevention of eating and weight-related problems through the internet. For example, our results suggested that providing social support in nutrition websites can be a venue to promote a healthy diet. We have discussed our results in more detail in the following lines.

Usage of Nutrition Websites

Being female, having higher levels of excessive exercise, and perceived online social support were positively associated with the usage of nutrition websites. The lack of association with eating disorder symptomatology and internalization of the beauty ideal suggests that the usage of nutrition websites is motivated more for health orientation rather than body image concerns or eating pathology. This orientation toward a healthier lifestyle could also explain the association between exercise and the usage of nutrition websites, although this association remains largely unexplored and further research is recommended. In any case, there is a global trend among some young women, including Czech women, to pursue a healthier lifestyle by improving their diet, nutrition, and fitness [29]. In this sense, nutrition-oriented websites can represent an important and easily accessible source of information. Dietary information on the internet, such as on blogs and social media, is found to be very useful for internet users given that they can find recipe ideas for cooking and social support to have a healthier diet and/or to pursue a healthier lifestyle [3,30]. However, the use of nutrition websites to pursue a healthier diet raises questions about the accuracy, quality, and impact of the health-related nutrition information. For instance, health communication among online users may be contaminated by inaccurate health information and/or by health beliefs originated from the misinterpretation of the information that they find on the internet [31,32]. Thus, more research attention should be paid to the cognitive processing, evaluation, and selection of health information made by internet users while navigating through healthy lifestyle websites. Moreover, considering the association between social support and nutrition websites, several initiatives can be undertaken to promote healthy eating through these kinds of websites. For example, governments and international organizations can design and promote online social communities that foster appropriate nutrition, rather than solely providing relevant nutrition information.

Usage of Weight Loss Websites

The usage of weight loss websites was positively associated with levels of exercise, internalization, being female, eating disorder symptomatology, and being overweight or obese. The usage of weight loss websites was also connected to age and online social support, although very weakly. Weight loss concerns are widespread among young women [33] and particularly among those with higher levels of internalization, those who are overweight, and those with disordered eating behavior [34,35]. Excessive exercise is frequently used for weight control purposes, and it is a common compensatory behavior among people with eating disorder symptomatology [36]. For overweight individuals, the internet is an easy-to-access source to find quick-fix weight loss solutions [17]. Therefore, taking all these individual factors together, it is possible that the underlying motivational factor for using these weight loss websites was body image concerns and particularly body weight concerns, which is a finding consistent with the literature [37]. Women who obtain weight loss information from the internet are more likely to exhibit unhealthy weight control behaviors [4]. Therefore, individuals' concerns about body image, eating, and weight could have a bidirectional association with their own usage of weight loss websites, and their selective exposure (deliberately or not) to this kind of media information may shape their own media effects [20]. Thus, the selective exposure to weight loss information and its effect on disordered eating is definitively a venue for future research exploring the role of individual differences (eg, information processing of health claims regarding weight loss).

Finally, it is important to note that internet users with certain characteristics such as body or weight concerns, eating disorder symptomatology, and excessive exercise are more likely to upload inspirational content (ie, to promote weight loss) as well as fitspirational content [21]. In turn, through psychological mechanisms, such as observational learning [19], viewers of this content may feel inspired to pursue the "thin ideal body" and to adopt unhealthy weight control behaviors. Thus, media literacy interventions aimed at promoting a critical examination of media messages regarding weight loss may serve as a useful public health initiative to ameliorate the potential harmful effects of these kinds of messages [38]. Moreover, these initiatives should promote digital literacy skills that can serve as a countermeasure against the internalization of the thin ideal, for example, educating an internet user on how to utilize digital tools that block weight loss advertising while surfing these websites.

Usage of Fitness Websites

The strongest association we found was between the levels of exercise and the frequency of fitness website usage, whereas its association with internalization was low and with age it was even lower. Following the selective exposure model [39], it could be suggested that individuals who engage in excessive exercise use fitness websites more frequently because the content on these websites is consistent with their beliefs and it can reinforce these beliefs [40]. In this sense, the role of internalization should be further explored. For instance, future research can examine how selective exposure to fitness websites among individuals with high levels of exercise may influence internalization as an enduring disposition or trait rather than as a state [35]. This approach would also contribute to further understanding the role of individual differences in the study of media effects [41].

Differences Among Individual Factors Associated With the Usage of Nutrition, Weight loss, and Fitness Websites

Concerning sociodemographic factors, we only found a weak association between age and the usage of weight loss and fitness websites, which could be due to the characteristics of our sampling procedure because it was focused on a younger population (aged 13 to 39 years). On the contrary, being female was associated with the higher usage of nutrition websites and weight loss websites, which seems a consistent finding in recent surveys [2,5]. However, there was no link between gender and the frequency of use of fitness websites. Previous studies on internet use regarding diet, weight, and physical activity have found secular trends by gender, suggesting changes over time regarding gender differences in internet use [2]. Moreover, there are seasonal and geographical variations regarding physical activity and dieting [42,43] as well as in the frequency of internet searches for fitness and weight loss information [44]. Therefore, certain sociodemographic characteristics associated with the usage of nutrition, weight loss, and fitness websites are manifested differently over time and place. Future research should further investigate this dynamic aspect of health-related internet use and the sociodemographic characteristics associated with it. This approach is useful in the design of tailored health communication campaigns directed toward specific segments of the population (eg, specific clusters of young women within social networks).

With regard to individual factors, excessive exercise was moderately associated with all 3 types of websites. Excessive exercise is not a unique characteristic of individuals with eating disorders or those practicing sports. Indeed, fitness activities that are apparently healthy may become problematic when they lead to excessive exercise patterns due to individual factors such as personality traits (eg, perfectionism) [45]. Therefore, this is an interesting finding suggesting that the variance in exercise motivation [46], coupled with other individual factors such as internalization, determines the variance in the usage of lifestyle-related websites. This finding highlights the importance of using integrative approaches in the study of health behavior such as integrating motivational theories regarding exercise [46], health behavior [47], and internet use [48].

Our findings also revealed that internalization was positively associated with the usage of weight loss and fitness websites but not with the usage of nutrition websites. A recent meta-analytic review found a strong relationship between internalization and the use of appearance-related features in social networking sites [49]. Similarly, internalization is associated with higher levels of compulsive exercise [50]. Therefore, beauty ideal internalization seems to be a better predictor of the consumption of appearance-based online media (eg, fitness and weight loss websites) rather than health-oriented online media such as nutrition information. Furthermore, research is recommended to confirm this hypothesis.

Finally, perceived online social support was not associated with the usage of fitness websites but it was associated with the usage of nutrition websites and weight loss websites, although this latter association was weak. As mentioned above, nutrition websites usually provide information regarding specific diets and healthy meals as well as social interactions that internet users find very useful for their eating and dietary needs and goals [3,30]. Nevertheless, it should be noted that privacy attitudes are important in online health communication, particularly regarding the self-disclosure of body weight and weight loss concerns. For instance, anonymity is easier online compared with offline peer-to-peer communication, and studies have found that this anonymity in the online context provides an opportunity for shared self-disclosure of eating and weight loss concerns among members of online communities and blog users [51]. Therefore, the results of our sample could suggest that social support is particularly relevant regarding the usage of nutrition websites and to some extent the usage of weight loss websites but most probably not relevant regarding the usage of fitness websites.

In conclusion, the frequency in the use of nutrition, weight loss, and fitness websites is associated with a different combination of individual characteristics. Public health initiatives should consider such individual differences in the design of online strategies for the prevention of eating and weight-related problems.

Limitations

Nevertheless, it is important to note that our study has some limitations. First, we explored the frequency of the usage of nutrition, weight loss, and fitness websites but not specific health behaviors related to these visits, such as the use of misinformation obtained from these websites. Thus, it would be beneficial to know more about how health-related information from these websites is actually used. In addition, the survey covered general patterns of visits to the 3 measured types of websites. However, it is probable that in some cases the visited websites could be defined by 2 or all 3 types. The data are also self-reported, a limitation which needs to be considered regarding the actual frequency of the visits (as compared with the recalled and reported one) as well as with eating disorder symptomatology, body weight, and height. Our sample is also limited in terms of generalizability, considering, for instance, that women were overrepresented and underweight individuals were underrepresented. Finally, the cross-sectional and correlational nature of our data precludes causal interpretations.

Although our study has limitations, it also has strengths. For instance, we asked participants about their frequency of use of specific websites rather than just asking them about online health information seeking. Moreover, this was a population-based survey that included diverse participants of both genders rather than a specific sample such as female college students. Finally, our results have important implications, particularly for the prevention of eating and weight-related problems such as eating disorders and obesity. Public health policies can be implemented to help with the personalization of Web content targeting individuals with a higher risk of developing eating and weight-related problems. For example, these policies can help to disseminate tailored public health messages (eg, about a healthy diet rather than weight loss), targeting specific users of lifestyle websites (eg, those with high internalization).

Acknowledgments

The authors acknowledge the support of the Czech Science Foundation (THINLINE-GA15-05696S) and the Faculty of Social Studies, Masaryk University, the Czech Republic.

Conflicts of Interest

None declared.

References

1. Fox S. Pew Research Center. 2005 May 17. Health information online URL: http://www.pewinternet.org/2005/05/17/health-information-online/ [accessed 2018-08-03] [WebCite Cache ID 71Ok7uUxi]
2. McCully SN, Don BP, Updegraff JA. Using the internet to help with diet, weight, and physical activity: results from the Health Information National Trends Survey (HINTS). J Med Internet Res 2013 Aug 01;15(8):e148 [FREE Full text] [doi: 10.2196/jmir.2612] [Medline: 23906945]
3. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: the perceived influence of social media on young adult health behaviors. Comput Human Behav 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]
4. Laz TH, Berenson AB. Association of web-based weight loss information use with weight reduction behaviors in adolescent women. J Adolesc Health 2011 Oct;49(4):446-448 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.01.020] [Medline: 21939880]
5. Pollard CM, Pulker CE, Meng X, Kerr DA, Scott JA. Who uses the internet as a source of nutrition and dietary information? An Australian population perspective. J Med Internet Res 2015 Aug 26;17(8):e209 [FREE Full text] [doi: 10.2196/jmir.4548] [Medline: 26310192]
6. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modelling and predicting eHealth usage in Europe: a multidimensional approach from an online survey of 13,000 European Union internet users. J Med Internet Res 2016 Dec 22;18(7):e188 [FREE Full text] [doi: 10.2196/jmir.5605] [Medline: 27450189]
7. Hoppmann C, Lay JC, Shayanfar S. Intraindividual variability in the context of adults' health behavior. In: Diehl M, Hooker K, Sliwinski MJ, editors. Handbook of Intraindividual Variability Across the Life-Span. New York: Routledge; Dec 16, 2014:216-228.
8. Dutta-Bergman M. Media use theory and internet use for health care. In: Murero M, Rice RE, editors. The Internet and Health Care: Theory, Research, and Practice. New York: Routledge; 2006:83-103.
9. Perloff RM. Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research. Sex Roles 2014 Dec;71(11-12):363-377. [doi: 10.1007/s11199-014-0384-6]
10. Fox S. Pew Research Center. 2011 Feb 28. Peer-to-peer health care URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [accessed 2016-12-28] [WebCite Cache ID 6n66SvlFg]
11. Vaterlaus JM, Jones RM, Patten EV, Cook JL. An exploratory study of time spent with interactive technology and body mass among young adults. Comput Human Behav 2015 Nov;52:107-114. [doi: 10.1016/j.chb.2015.05.035]
12. Wangberg SC, Sørensen T, Andreassen HK. Using the Internet to support exercise and diet: a stratified Norwegian survey. Med 2 0 2015 Aug 26;4(2):e3 [FREE Full text] [doi: 10.2196/med20.4116] [Medline: 26310277]
13. Carrotte ER, Vella AM, Lim MS. Predictors of "Liking" three types of health and fitness-related content on social media: a cross-sectional study. J Med Internet Res 2015 Aug 21;17(8):e205 [FREE Full text] [doi: 10.2196/jmir.4803] [Medline: 26297689]
14. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
15. Rodgers RF, Skowron S, Chabrol H. Disordered eating and group membership among members of a pro-anorexic online community. Eur Eat Disord Rev 2012 Jan;20(1):9-12. [doi: 10.1002/erv.1096] [Medline: 21305675]

16. Puhl RM, Peterson JL. The nature, consequences,public health implications of obesity stigma. In: Corrigan PW, editor. The Stigma of Disease and Disability: Understanding Causes and Overcoming Injustices. Washington, DC: American Psychological Association; 2014:183-203.
17. Lewis S, Thomas SL, Blood RW, Castle D, Hyde J, Komesaroff PA. 'I'm searching for solutions': why are obese individuals turning to the internet for help and support with 'being fat'? Health Expect 2011 Dec;14(4):339-350. [doi: 10.1111/j.1369-7625.2010.00644.x] [Medline: 21199200]
18. Turner JS. Negotiating a media effects model: addendums and adjustments to Perloff's framework for social media's impact on body image concerns. Sex Roles 2014 Nov 9;71(11-12):393-406. [doi: 10.1007/s11199-014-0431-3]
19. Levine MP, Harrison K. Effects of media on eating disorders and body image. In: Bryant J, Oliver MB, editors. Media Effects: Advances in Theory and Research. 3rd edition. New York: Routledge; 2009:490-516.
20. Valkenburg PM, Peter J, Walther JB. Media effects: theory and research. Annu Rev Psychol 2016;67:315-338. [doi: 10.1146/annurev-psych-122414-033608] [Medline: 26331344]
21. Holland G, Tiggemann M. "Strong beats skinny every time": disordered eating and compulsive exercise in women who post fitspiration on Instagram. Int J Eat Disord 2017 Jan;50(1):76-79. [doi: 10.1002/eat.22559] [Medline: 27302867]
22. Hefner V, Dorros SM, Jourdain N, Liu C, Tortomasi A, Greene MP, et al. Mobile exercising and tweeting the pounds away: the use of digital applications and microblogging and their association with disordered eating and compulsive exercise. Cogent Soc Sci 2016 Apr 28;2(1):-. [doi: 10.1080/23311886.2016.1176304]
23. Forbush KT, Wildes JE, Pollack LO, Dunbar D, Luo J, Patterson K, et al. Development and validation of the Eating Pathology Symptoms Inventory (EPSI). Psychol Assess 2013 Sep;25(3):859-878. [doi: 10.1037/a0032639] [Medline: 23815116]
24. Hill LS, Reid F, Morgan JF, Lacey JH. SCOFF, the development of an eating disorder screening questionnaire. Int J Eat Disord 2010 May;43(4):344-351. [doi: 10.1002/eat.20679] [Medline: 19343793]
25. World Health Organization. Obesity: Preventing and managing the global epidemic. Geneva: World Health Organization; 2000.
26. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes 2012 Aug;7(4):284-294. [doi: 10.1111/j.2047-6310.2012.00064.x] [Medline: 22715120]
27. Cusumano DL, Thompson JK. Media influence and body image in 8-11-year-old boys and girls: a preliminary report on the multidimensional media influence scale. Int J Eat Disord 2001 Jan;29(1):37-44. [Medline: 11135331]
28. Graham AL, Papandonatos GD, Kang H, Moreno JL, Abrams DB. Development and validation of the online social support for smokers scale. J Med Internet Res 2011 Sep 28;13(3):e69 [FREE Full text] [doi: 10.2196/jmir.1801] [Medline: 21955465]
29. Euromonitor International. London: Euromonitor International Ltd; 2015 Sep. Young women's health: global attitudes towards health, fitness and wellbeing among the under 30s and market impact. Book URL: https://www.euromonitor.com/young-womens-health-global-attitudes-towards-health-fitness-and-wellbeing-among-the-under-30s-and-market-impact/report [WebCite Cache ID 76RsIL7Tw]
30. Bissonnette-Maheux V, Provencher V, Lapointe A, Dugrenier M, Dumas AA, Pluye P, et al. Exploring women's beliefs and perceptions about healthy eating blogs: a qualitative study. J Med Internet Res 2015 Apr 8;17(4):e87 [FREE Full text] [doi: 10.2196/jmir.3504] [Medline: 25858777]
31. Bonnar-Kidd KK, Black DR, Mattson M, Coster D. Online physical activity information: will typical users find quality information? Health Commun 2009 Mar;24(2):165-175. [doi: 10.1080/10410230802676763] [Medline: 19280460]
32. Mitu B. Health in the digital era: searching health information online. In: Marinescu V, Mitu B, editors. The Power of the Media in Health Communication. New York: Routledge; 2016:146-156.
33. Julia C, Péneau S, Andreeva VA, Méjean C, Fezeu L, Galan P, et al. Weight-loss strategies used by the general population: how are they perceived? PLoS One 2014 May 22;9(5):e97834 [FREE Full text] [doi: 10.1371/journal.pone.0097834] [Medline: 24852440]
34. Neumark-Sztainer DR. Higher weight status and restrictive eating disorders: an overlooked concern. J Adolesc Health 2015 Jan;56(1):1-2. [doi: 10.1016/j.jadohealth.2014.10.261] [Medline: 25530600]
35. Karazsia BT, van Dulmen HM, Wong K, Crowther JH. Thinking meta-theoretically about the role of internalization in the development of body dissatisfaction and body change behaviors. Body Image 2013 Sep;10(4):433-441. [doi: 10.1016/j.bodyim.2013.06.005] [Medline: 23871195]
36. Fernandez-del-Valle M. Screening tools for excessive exercise in the active female. In: Robert-McComb JJ, Norman RL, Zumwalt M, editors. The Active Female: Health Issues Throughout the Lifespan. New York: Springer; 2014:373-387.
37. Rodgers RF, Melioli T. The relationship between body image concerns, eating disorders and internet use, Part I: a review of empirical support. Adolesc Res Rev 2016 Jun;1(2):95-119. [doi: 10.1007/s40894-015-0016-6]
38. Bergsma L, Ferris E. The impact of health-promoting media-literacy education on nutrition and diet behavior. In: Preedy VR, Watson RR, Martin CR, editors. Handbook of Behavior, Food and Nutrition. New York: Springer; 2011:3391-3411.
39. Knobloch-Westerwick S, Sarge MA. Impacts of exemplification and efficacy as characteristics of an online weight-loss message on selective exposure and subsequent weight-loss behavior. Commun Res 2015 Jun;42(4):547-568. [doi: 10.1177/0093650213478440]

40. 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 01;17(3):281-303. [doi: 10.1111/j.1468-2885.2007.00296.x]
41. Oliver MB, Krakowiak KM. Individual differences in media effects. In: Bryant J, Zillmann D, editors. Media Effects: Advances in Theory and Research. 3rd edition. New York: Routledge; 2009:517-531.
42. Gracia-Marco L, Ortega FB, Ruiz JR, Williams CA, Hagströmer M, Manios Y, Helena Study Group. Seasonal variation in physical activity and sedentary time in different European regions. The HELENA study. J Sports Sci 2013;31(16):1831-1840. [doi: 10.1080/02640414.2013.803595] [Medline: 24050788]
43. Markey PM, Markey CN. Annual variation in internet keyword searches: linking dieting interest to obesity and negative health outcomes. J Health Psychol 2013 Jul;18(7):875-886. [doi: 10.1177/1359105312445080] [Medline: 22992586]
44. Madden KM. The seasonal periodicity of healthy contemplations about exercise and weight loss: ecological correlational study. JMIR Public Health Surveill 2017 Dec 13;3(4):e92 [FREE Full text] [doi: 10.2196/publichealth.7794] [Medline: 29237582]
45. Cook B, Hausenblas H, Freimuth M. Exercise addiction and compulsive exercising: relationship to eating disorders, substance use disorders, and addictive disorders. In: Brewerton T, Dennis AB, editors. Eating Disorders, Addictions and Substance Use Disorders: Research, Clinical and Treatment Perspectives. Berlin: Springer-Verlag; 2014:127-144.
46. Hagger MS. Advances in motivation in exercise and physical activity. In: Ryan RM, editor. The Oxford Handbook of Human Motivation. New York: Oxford University Press; 2012:479-504.
47. Dunsmore S, Goodson P. Motivation for healthy behavior: a review of health promotion research. Am J Health Educ 2013 Jan 23;37(3):170-183. [doi: 10.1080/19325037.2006.10598897]
48. Amiel T, Sargent SL. Individual differences in internet usage motives. Comput Hum Behav 2004 Nov;20(6):711-726. [doi: 10.1016/j.chb.2004.09.002]
49. Mingoia J, Hutchinson AD, Wilson C, Gleaves DH. The relationship between social networking site use and the internalization of a thin ideal in females: a meta-analytic review. Front Psychol 2017 Aug 07;8:1351 [FREE Full text] [doi: 10.3389/fpsyg.2017.01351] [Medline: 28824519]
50. Homan K. Athletic-ideal and thin-ideal internalization as prospective predictors of body dissatisfaction, dieting, and compulsive exercise. Body Image 2010 Jun;7(3):240-245. [doi: 10.1016/j.bodyim.2010.02.004] [Medline: 20226748]
51. Leggatt-Cook C, Chamberlain K. Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere. Sociol Health Illn 2012 Sep;34(7):963-977 [FREE Full text] [doi: 10.1111/j.1467-9566.2011.01435.x] [Medline: 22150320]

Abbreviations

BMI: body mass index

Edited by G Eysenbach; submitted 21.02.18; peer-reviewed by PH Lin, R Newton; comments to author 03.08.18; revised version received 20.11.18; accepted 26.01.19; published 04.04.19.

Please cite as:
Almenara CA, Machackova H, Smahel D
Sociodemographic, Attitudinal, and Behavioral Correlates of Using Nutrition, Weight Loss, and Fitness Websites: An Online Survey
J Med Internet Res 2019;21(4):e10189
URL: https://www.jmir.org/2019/4/e10189/
doi:10.2196/10189
PMID:30946018

Original Paper

Web-Based Health Information Technology: Access Among Latinos Varies by Subgroup Affiliation

Mariaelena Gonzalez[1,2*], PhD; Ashley Sanders-Jackson[3*], PhD; Tashelle Wright[1], MSPH

[1]Department of Public Health, University of California - Merced, Merced, CA, United States

[2]Health Science Research Institute, University of California - Merced, Merced, CA, United States

[3]Department of Advertising and Public Relations, College of Communication Arts and Sciences, Michigan State University, East Lansing, MI, United States

*these authors contributed equally

Corresponding Author:
Mariaelena Gonzalez, PhD
Department of Public Health
University of California - Merced
5200 N Lake Road
Merced, CA,
United States
Phone: 1 650 283 8307
Email: mgonzalez82@ucmerced.edu

Abstract

Background: There are significant health technology gaps between Latinos and non-Hispanic whites and between first- and second-generation Latinos.

Objective: This study aimed to examine disparities in Web-based health information–seeking behavior (HISB) and patient portal use among Latinos, taking into account nativity and subethnic affiliation.

Methods: We analyzed US-born, non-Hispanic whites and Latinos adults (N=49,259) and adult internet users (N=36,214) in the 2015 to 2016 National Health Interview Survey using a binary logistic regression controlling for individual difference level variables. Outcomes were internet use, HISB (health information-seeking online and using a chat group for health information), and patient portal use (using a computer to schedule an appointment, filling a prescription, and communicating with a provider).

Results: We found that US-born Mexicans (odds ratio [OR] 0.81, 95% CI 0.66-0.99), foreign-born Mexicans (OR 0.35, 95% CI 0.29-0.42), foreign-born Puerto Ricans (OR 0.62, 95% CI 0.44-0.87), foreign-born Central and South Americans (OR 0.42, 95% CI 0.33-0.53), and foreign-born other Latinos (OR 0.34, 95% CI 0.24-0.49) had lower odds of using the internet than US-born non-Hispanic whites. The relationship between subgroup affiliation and Web-based HISB varied by type of technology. US-born Mexicans (OR 0.77, 95% CI 0.66-0.9), foreign-born Mexicans (OR 0.51, 95% CI 0.43-0.61), foreign-born Central and South Americans (OR 0.53, 95% CI 0.43-0.64), and foreign-born other Latinos (OR 0.56, 95% CI 0.4-0.79) had lower odds of looking up health information online than US-born non-Hispanic whites. Controlling for age, sex, education, income to federal poverty level, and region, foreign-born Central and South Americans (OR 0.61, 95% CI 0.41-0.92) and foreign-born other Latinos (OR 0.26, 95% CI 0.1-0.68) had lower odds of filling a prescription using a computer than US-born non-Hispanic whites. Foreign-born Mexicans (OR 0.51, 95% CI 0.36-0.72) and foreign-born Central and South Americans (OR 0.7, 95% CI 0.5-0.99) have lower odds of emailing a health care provider than US-born non-Hispanic whites. Posthoc analyses were conducted among Mexican-Americans to see if age was significant in predicting Web-based HISB or other patient portal use. We found individuals aged 18 to 30 years had higher odds of using the internet (OR 3.46, 95% CI 2.61-4.59) and lower odds of looking up health information online (OR 0.75, 95% CI 0.58-0.96). A posthoc analysis was conducted among Mexican-Americans to see if nativity predicted Web-based HISB and patient portal use. We found that US-born individuals had higher odds (OR 52.9, 95% CI 1.2-1.93) of looking up health information online compared with foreign-born individuals.

Conclusions: We found Latino subgroups do not use health information channels equally, and attempts to target Latinos should take ethnicity and nativity into account.

(J Med Internet Res 2019;21(4):e10389) doi:10.2196/10389

KEYWORDS

information seeking behavior; health information technology; Hispanic Americans; minority health

Introduction

Background

The use of Web-based health information technology (HIT) is spreading [1]. For example, encouraged by the Health Information Technology for Economic and Clinical Health Act, in 2012, at least half of all health care providers had adopted the use of patient portals [2]. Patient portals have been shown to have a wide variety of benefits, including the management of chronic disease; however, Latinos are also less likely to adopt patient portals [3,4]. Previous research has suggested disparities in internet access, online health information–seeking behavior (HISB) between non-Hispanic (NH) whites and Latinos, and between US- and foreign-born Latinos [5-11]. With the increasing reliance on patient portals [4], it is important to understand where disparities in specific types of online HISB and patient portal use occur for Latinos at the national level.

Although Latinos have historically lagged behind NH whites in access to the internet, there have been significant gains in internet access among Latinos between 2012 and 2015, particularly among Spanish speakers [8]. Despite this, only 46% of Hispanics access the internet through broadband, and a 2017 Pew Research Report stated that 22% of Hispanics did not have broadband but owned a smartphone [8,12]. These differential patterns of access and device use are disparities, and disparities in the use of Web-based HIT, such as patient portals, should also be investigated.

There are significant information knowledge gaps between Latinos and NH whites when it comes to Web-based patient portals [4,13-15]. In a systematic review of patient portal literature, although patient portals may improve outcomes, particularly in conjunction with a clinical intervention, for some patients with chronic conditions, racial and ethnic minorities may not receive the same benefit as whites [15]. Another systematic review found that ethnicity can play a role in lack of adoption of patient portals [4]. Indeed, in a California sample of elderly adults, Latinos were less likely to use patient portals than NH whites, although Latinos were not broken down into subgroups in this analysis [16]. Latinos also have lower rates of activating and logging into patient portals than NH whites [17,18].

Although some state- and local-level examinations of online HISB and the use of patient portals exist [9,19-22], and several national level studies [1,11], these studies do not control for other factors such as Latino subethnicity or nativity that might be important cultural variables that could contribute to or mitigate disparities.

When examining Latinos in the United States, nativity and ethnicity are important determinants of health and media use. Although Latinos are the largest ethnic group in the United States [23], there are significant demographic differences between US- and foreign-born individuals. On average, US-born Latinos, who make up over 60% of the Latino population in the United States, are younger and consume more English-language media [8,24]. US-born Latinos engage in more online HISB and have higher levels of confidence in their ability to fill out Web-based forms than foreign-born Latinos [9]. Although researchers commonly lump Latino subethnicities together, there are significant health and behavioral differences within and between the various subethnicities [25-28]. For example, research suggests that Puerto Ricans are more likely to be current smokers and to have ever smoked compared with Dominicans, Colombians, and Ecuadorian individuals [27]. Puerto Ricans also reportedly have higher rates of depression and suicide attempts compared with Mexican and Cuban Americans [26], Cuban Americans have higher systolic blood pressure and higher cholesterol than both Mexican Americans and Puerto Ricans [25], and perceptions of cancer risk factors vary across Latino ethnic subgroups [28]. A study in 2007, which examined internet use among Latinos by subethnicity, found that in the United States, internet use varies among ethnic groups, with South American origin Latinos having a higher than average rate of internet use and Mexicans and Central Americans having lower rates of internet use [29], which indicates that ethnicity is an important factor when examining internet-related behavior among Latinos. However, we did not find other studies on Latino internet use that disaggregate Latinos in this way. Due to the US or foreign-born (nativity) and ethnic differences among Latinos, it is important to take into account nativity and ethnicity, if possible, when examining disparities between NH whites and Latinos.

Objective

The purpose of this study was to examine disparities in online HISB and patient portal use among NH whites, US-born Latinos, and foreign-born Latinos, taking into account subethnic affiliation and disaggregating subethnicities or regional groupings when possible. By understanding where disparities exist, we can effectively target interventions to increase online health literacy within the largest linguistically and culturally sensitive ethnic group in the United States.

Methods

Data

We combined 2 years (2015 and 2016) of the adult sample file of the National Health Interview Survey (NHIS), which is a nationally representative sample of the adult US civilian population. We examined all US-born whites and Latinos (N=49,251) and only those individuals who reported using the internet (N=36,214).

Primary Outcomes

Respondents were asked if in the last 12 months they had used the internet or used email. In addition, they were asked if in the last year they had engaged in online HISB and used a computer to (1) “[l]ook up health information on the Internet,” [30] or (2) “[u]se online chat groups to learn about health topics” [30]. They were also asked if they had used patient portals by using a computer to (3) “[f]ill a prescription,” [30] (4) “[s]chedule an appointment with a health care provider,” [30] or (5) “[c]ommunicate with a health care provider by email” [30]. Each of these binary items was separately used as a primary outcome in our analyses. If respondents reported using the internet, email, engaging in online HISB or online patient portals in the last 12 months, they were classified as having used the internet in the last 12 months.

Latino Subgroup Affiliation by Nativity

There are differences between US- and foreign-born Hispanics with regard to the median age and use of Spanish or English within the population [9,24,31,32]. There are differences between Hispanic subgroups and region, which may play a role in health care access [25-28]. Furthermore, different countries of origin may be associated with cultural differences that impact health access. In addition, studies have shown significant behavioral and health differences among various Hispanic ethnic groups, including around internet use [29,31-33]. Although combining 2 years of the NHIS did not yield a large enough sample to disaggregate all Latino ethnic subgroups in our model, we were able to obtain a large enough sample to divide Hispanics into the following subgroups: Mexican and Mexican-Americans (Mexicans), Puerto Ricans, Central and South American, and other Latino (Cuban, Dominican, Other Latin American affiliation). Although some subgroups we used contain large regions or disparate groups, we combined these groups based on Centers for Disease Control and Prevention’s classifications and because of sample size issues. Individuals who identified as *other Spanish* or *multiple Hispanic* were excluded as we could not specifically identify their ethnic or regional origin. We further divided each of the 4 Latino subgroups into US- and foreign-born individuals. Our comparison group for all Latino categories was US-born NH whites, as it is possible that foreign-born NH whites have different health outcomes and media use patterns than US-born individuals.

Other Individual-Level Variables

We controlled for age (18-30 years, 31-54 years, and 55 years and above), gender (male, female), education (below a high school education, completed high school or general education development, some college, and completed a bachelor’s or higher), poverty level (0%-99% of the federal poverty level [FPL], 100%-199% of the FPL, and 200% or greater of the FPL), marital status (married or partnered, unpartnered), and whether or not the respondent was insured. We also controlled for the individuals’ occupational group, classifying individuals who were employed by their 4-category occupational group status (white collar, blue collar, farm and forestry, and service) and those who were not in the labor force (individuals who reported that they were looking for work, or not working at a job or business, and not looking for work). We could not use this 5-category occupational and employment designations in our posthoc statistical analyses as 138 of the 170 Mexican-descent individuals in the farm, or fishing or forestry category were foreign-born, which means there was little variation in this category among Mexican-identified individuals, so we combined this category with the blue-collar category in the posthoc analyses. In addition, we controlled for region of the United States (North, Midwest, South, and West) and whether or not the respondent reported using the internet at least once a day. To minimize collinearity, we did not control for English proficiency, as we found it significantly negatively correlated with being US-born among our sample of US-born whites and Latinos (correlation=−0.66).

Statistical Analysis

We ran multivariable binary logistic regressions to test the relationship between ethnicity and nativity and internet use. We then used a binary logistic regression to test the relationship between ethnicity and nativity and our 2 HISB and 3 patient portal outcomes among internet users. The sample was weighted according to survey directions. Due to the small number of individuals among some of the Latino ethnic and nativity groups, we do not report groups where too few individuals use the technology in question to draw a statistical conclusion.

Results

Demographics

The majority (38,256/49,259, 82.6%; 95% CI 82.1-83.2) of respondents reported using the internet (Table 1), and of 65.05% (25,209/38,256; 95% CI 64.3-65.8) respondents reported looking up health information on the internet; however, less than 13% of respondents reported using a computer to engage in any of the surveyed Web-based patient portal channels (Table 2). Among internet users, Mexican-identified individuals (Mexicans) were the largest US- and foreign-born subgroups (2337/38,256, 6.24%; 95% CI 5.8-6.8 and 1358/38,256, 4.1%; 95% CI 3.8-4.4, respectively). The next largest US-born subgroup was Puerto Ricans (414/38,256, 1.2%; 95% CI 0.97-1.24). There were more foreign-born Central and South Americans (739/38,256) and other Latinos (307/38,256) than US-born individuals within these subgroupings (246/38,256 and 139/38,256, respectively).

Table 1. Characteristics of US-born non-Hispanic whites and Latinos (National Health Interview Survey, 2015-16). N=49,259.

Variables	Statistics, n (%)	95% CI
Dependent variables		
Use the internet	38,256 (82.61)	82.05-83.15
Independent variables		
Ethnicity by nativity NH[a] whites and Latino subgroup affiliation		
US-born NH White	40,732 (80.36)	79.34-81.34
US-born Mexican	2585 (6.06)	5.56-6.79
Foreign-born Mexican	2718 (6.28)	5.8-6.79
US-born Puerto Rican	489 (1.09)	0.97-1.24
Foreign-born Puerto Rican	476 (0.99)	086-1.14
US-born Central or South American	268 (0.71)	0.59-0.85
Foreign-born Central or South American	1195 (2.86)	2.59-3.16
US-born other Latino	157 (0.42)	0.33-0.52
Foreign-born other Latino	631 (1.22)	1.04-1.43
Age, years		
18-30	8837 (22.21)	21.58-22.85
31-54	18,386 (40.29)	39.65-40.94
>55	22,036 (37.5)	36.81-38.19
Sex		
Female	26,633 (51.18)	50.57-51.79
Male	22,626 (48.82)	48.21-49.43
Education		
Below high school	6152 (12.37)	11.87-12.88
High school or general education development	12,258 (24.92)	24.33-25.52
Some college	15,633 (31.53)	30.91-32.16
Bachelor’s or above (reference)	15,072 (31.18)	30.41-31.96
Income to federal poverty level (FPL) ratio		
Poor (income to FPL ratio 0-0.99)	6266 (10.66)	10.22-11.12
Near poor (income to FPL ratio 1-1.99)	8843 (17.32)	16.8-17.85
Above poor (income to FPL ratio >2)	31,788 (72.02)	71.31-72.73
Married or partnered		
Married or partnered	26,135 (63.48)	62.83-64.13
Single or widowed	23,040 (36.52)	35.87-37.17
Occupational group		
White collar	28,234 (58.12)	57.39-58.85
Service	7615 (15.37)	14.89-15.86
Farm and forestry	475 (0.86)	0.73-1.01
Blue Collar	10,168 (21.03)	20.46-21.61
Not in labor force	2138 (4.63)	4.32-4.96
Health insurance		
Covered	44,471 (90.16)	89.71-90.6
Not covered	4613 (9.84)	9.4-10.29

Variables	Statistics, n (%)	95% CI
Region		
Northeast	8473 (17.49)	16.67-18.35
Midwest	11,801 (24.51)	23.66-25.38
South	15,617 (34.44)	33.27-35.64
West	13,368 (23.55)	22.45-24.69

[a] NH: non-Hispanic.

Internet Use

The percentage of internet users who use each technology varies by ethnicity and nativity (Tables 3-5). A total of 85.6% (32,735/40,732; 95% CI 85.1-86.1) of US-born NH whites report using the internet, whereas 53.76% (1358/2037; 95% CI 51.1-56.4) of US-born Mexicans reported using the internet. Foreign-born Mexicans (633/2037) and foreign-born Central and South Americans (370/739) and foreign-born other Latinos (145/307) had less than 50% of the population report looking up health information online.

We found that US-born Mexicans (OR 0.81, 95% CI 0.66-0.99), foreign-born Mexicans (OR 0.35, 95% CI 0.29-0.42), foreign-born Puerto Ricans (OR 0.62, 95% CI 0.44-0.87), foreign-born Central and South Americans (OR 0.42, 95% CI 0.33-0.53), and foreign-born other Latinos (OR 0.34, 95% CI 0.24-0.49) had lower odds of using the internet than US-born NH whites (Table 6).

Online Health Information–Seeking Behavior

The relationship between subgroup affiliation and online HISB varied by type of technology. US-born Mexicans (OR 0.77, 95% CI 0.66-0.90), foreign-born Mexicans (OR 0.51, 95% CI 0.43-0.61), foreign-born Central and South Americans (OR 0.53, 95% CI 0.43-0.64), and foreign-born other Latinos (OR 0.56, 95% CI 0.40-0.79) had lower odds of looking up health information online than US-born NH whites (Table 6). Foreign-born Central and South Americans (OR 1.71, 95% CI 1.11-2.62) had higher odds of using a chat group to discuss health information than US-born NH whites, whereas there was no significant difference between all other groups and NH whites regarding the use of chat groups to discuss health information.

Patient Portal Use

Controlling for age, sex, education, income to FPL, and region, foreign-born Central and South Americans (OR 0.61, 95% CI 0.41-0.92) had lower odds of filling a prescription using a computer than US-born NH whites (Table 7). There was no significant difference using a computer to schedule medical appointments between US-born NH whites and the various Latino ethnic and nativity groups. Foreign-born Mexicans (OR 0.51, 95% CI 0.36-0.72) and foreign-born Central and South Americans (OR 0.51, 95% CI 0.36-0.72) have lower odds of emailing a health care provider than US-born NH whites.

Table 2. Characteristics of US-born non-Hispanic whites and Latino internet users (National Health Interview Survey, 2015-16). N=38,256.

Variables	Statistics, n (%)	95% CI
Dependent variables		
Use the internet	—[a]	—
Online health information–seeking behavior		
Look up health information on the internet	25,209 (65.05)	64.3-65.8
Use online chat groups to learn about health topics	1634 (4.32)	4.06-4.59
Web-based health information technology		
Used a computer to fill a prescription	4564 (11.91)	11.43-12.41
Used a computer to schedule an appointment with a health care provider	4834 (13.43)	12.87-14.01
Used a computer to communicate with a health care provider by email	5653 (15.2)	14.6-15.82
Independent variables		
Ethnicity by nativity NH[b] whites and Latino subgroup affiliation		
US-born NH white	32,735 (83.28)	82.41-84.11
US-born Mexican	2337 (6.24)	5.75-6.76
Foreign-born Mexican	1358 (4.08)	3.76-4.44
US-born Puerto Rican	414 (1.18)	1.03-1.35
Foreign-born Puerto Rican	276 (0.78)	0.66-0.93
US-born Central and South American	246 (0.82)	0.68-0.99
Foreign-born Central and South American	739 (2.33)	2.07-2.63
US-born other Latino	139 (0.46)	0.37-0.59
Foreign-born other Latino	307 (0.8)	0.67-0.96
Age, years		
18-30	8191 (25.37)	24.64-26.12
31-54	15,772 (42.97)	42.26-43.69
>55	14,293 (31.66)	30.94-32.39
Sex		
Female	20,790 (51.51)	50.83-52.19
Male	17,466 (48.49)	47.81-49.19
Education		
Below high-school	2488 (7.09)	6.71-70.5
High-school or general education development	8253 (22.20)	21.59-22.83
Some college	13,348 (34.26)	33.54-34.98
Bachelor’s or above (reference)	14,106 (36.45)	35.58-37.32
Income to federal poverty level (FPL) ratio		
Poor (income to FPL ratio 0-0.99)	3899 (8.41)	7.99-8.86
Near poor (income to FPL ratio 1-1.99)	5630 (14.34)	13.81-14.88
Above poor (income to FPL ratio >2)	27,237 (77.25)	76.55-77.93
Married or partnered		
Married or partnered	21,536 (65.36)	64.63-66.09
Single or widowed	16,664 (34.64)	33.91-35.37
Occupational group		
White collar	24,659 (64.45)	63.71-65.18

Variables	Statistics, n (%)	95% CI
Service	5403 (14.28)	13.76-14.81
Farm and forestry	228 (0.51)	0.42-0.6
Blue collar	6452 (17.48)	16.94-18.04
Not in labor force	1028 (3.29)	3.01-3.59
Use the internet once a day or more often		
Use the internet once a day or more often	28,591 (78.32)	77.41-79.21
Use the internet less than once a day	7804 (21.68)	20.79-22.59
Health insurance		
Covered	34,801 (91.21)	90.78-91.62
Not covered	3319 (8.79)	8.38-9.22
Region		
Northeast	6576 (17.68)	16.8-18.59
Midwest	9273 (24.9)	23.97-25.86
South	11,653 (33.27)	32.07-34.5
West	10,772 (24.15)	23.02-25.31

[a] —: not applicable.

[b] NH: non-Hispanic.

Table 3. Percent of individuals in each ethnic-native group who use the internet and look up health information online (National Health Interview Survey, 2015-16).

Ethnicity by nativity	Population: use the internet		Population: look up health information	
	n (%)	95% CI	n (%)	95% CI
US-born non-Hispanic whites	32,735 (85.6)	85.1-86.1	22,167 (67.3)	66.6-68.1
US-born Mexican	2037 (85.0)	82.7-87.1	1228 (58.2)	55.0-61.4
Foreign-born Mexican	1358 (53.8)	51.1-56.4	633 (45.1)	41.6-48.7
US-born Puerto Rican	414 (88.9)	84.0-92.4	247 (58.2)	52.1-64.0
Foreign-born Puerto Rican	276 (65.2)	60.2-70.0	155 (56.3)	48.2-64.0
US-born Central and South American	246 (95.9)	92.5-97.7	171 (69.1)	60.6-76.5
Foreign-born Central and South American	739 (67.3)	63.6-70.8	370 (48.8)	44.1-53.5
US-born other Latino	139 (91.4)	83.7-95.6	90 (61.7)	51.1-71.3
Foreign-born other Latino	307 (55.4)	50.0-60.7	145 (45.8)	38.8-52.9

Table 4. Percent of individuals in each ethnic-native group who are internet users and use a chat group or fill a prescription online (National Health Interview Survey, 2015-16).

Ethnicity by nativity	Population: use chat group		Population: fill prescription	
	n (%)	95% CI	n (%)	95% CI
US-born whites	1349 (4.2)	4.0-4.5	4170 (12.9)	12.3-13.4
US-born Mexican	93 (4.1)	3.1-5.4	184 (9.2)	7.7-11.2
Foreign-born Mexican	73 (4.7)	3.6-6.2	62 (5.0)	3.7-6.7
US-born Puerto Rican	21 (3.4)	2.0-5.8	37 (8.2)	5.0-13.2
Foreign-born Puerto Rican	15 (6.4)	3.2-12.5	19 (6.0)	3.5-9.9
US-born Central and South American	12 (3.5)	1.7-6.9	30 (8.9)	5.5-14.0
Foreign-born Central and South American	52 (8.0)	5.6-11.3	46 (6.5)	4.5-9.2
US-born other Latino	3 (2.1)	0.5-8.4	8 (4.3)	1.9-9.4
Foreign-born other Latino	16 (3.9)	2.0-7.5	7 (2.3)	1.0-5.3

Table 5. Percent of individuals in each ethnic-native group who use the internet and look up health information online (National Health Interview Survey 2015-16).

Ethnicity by nativity	Population: schedule medical appointment		Population: email providers	
	n (%)	95% CI	n (%)	95% CI
US-born whites	4278 (14.1)	13.5-14.7	5177 (16.5)	15.9-17.2
US-born Mexican	247 (12.7)	10.8-14.8	209 (10.9)	9.1-13.1
Foreign-born Mexican	89 (6.8)	5.3-8.7	66 (4.8)	3.5-6.5
US-born Puerto Rican	47 (12.2)	8.4-17.2	46 (9.3)	6.3-13.5
Foreign-born Puerto Rican	25 (7.9)	5.0-12.3	17 (6.0)	3.6-9.9
US-born Central and South American	47 (17.1)	11.7-24.3	40 (16.2)	10.5-24.2
Foreign-born Central and South American	71 (9.5)	7.2-12.4	71 (9.4)	7.0-12.5
US-born other Latinos	13 (10.4)	5.8-17.7	11 (5.9)	3.0-11.3
Foreign-born other Latinos	17 (5.3)	3.1-8.9	16 (5.6)	2.9-10.6

Table 6. Correlation between (1) Latino subgroup affiliation and internet use and (2) Latino subgroup affiliation and online health information seeking behavior among internet users, controlling for age, sex, education, income to federal poverty level, marital or partnered status, occupation, insurance status, daily internet use, and region using the National Health Interview Survey 2015-16.

Variables[a]	Dependent variable: use internet (n=45,449), OR[b] (95% CI)	Internet users	
		Dependent variable: look up health information (n=34,467), OR (95% CI)	Dependent variable: use chat group (n=34,484), OR (95% CI)
US-born white	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
US-born Mexican	0.81 (0.66-0.99)[c]	0.77 (0.66-0.9)[c]	1.02 (0.74-1.42)
US-born Puerto Rican	1.2 (0.78-1.84)	0.75 (0.56-1.01)	0.96 (0.54-1.7)
US-born Central and South American	1.44 (0.75-2.74)	1.16 (0.76-1.77)	0.92 (0.44-1.91)[d]
US-born other Latino	1 (0.41-2.47)	0.83 (0.49-1.4)	0.5 (0.11-2.18)[d]
Foreign-born Mexican	0.35 (0.29-0.42)[c]	0.51 (0.43-0.61)[c]	1.28 (0.9-1.82)
Foreign-born Puerto Rican	0.62 (0.44-0.87)[c]	0.82 (0.57-1.17)	1.27 (0.54-2.99)
Foreign-born Central and South American	0.42 (0.33-0.53)[c]	0.53 (0.43-0.64)[c]	1.71 (1.11-2.62)[c]
Foreign-born other Latino	0.34 (0.24-0.49)[c]	0.56 (0.4-0.79)[c]	0.92 (0.43-1.97)[d]

[a] Control variables not listed include age, sex, education, income to federal poverty level ratio, occupational group, insurance status, daily internet use, and region.

[b] OR: odds ratio.

[c] Indicates values are significant.

[d] Too few individuals use the technology in question to draw a statistical conclusion.

Table 7. Correlation between Latino subgroup affiliation and online patient portal use among internet users controlling for age, sex, education, income to federal poverty level, marital or partnered status, occupation, insurance status, daily internet use, and region using the National Health Interview Survey 2015-16. N=46,480.

Variables[a]	Use of health information technology		
	Dependent variable: fill prescription (n=34,483), OR[b] (95% CI)	Dependent variable: schedule medical appointment (n=34,477), OR (95% CI)	Dependent variable: email provider (n=34,480), OR (95% CI)
US-born non-Hispanic white	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
US-born Mexican	1.01 (0.81-1.26)	1.01 (0.83-1.23)	0.85 (0.68-1.06)
US-born Puerto Rican	0.94 (0.53-1.68)	1.12 (0.71-1.77)	0.8 (0.5-1.27)
US-born Central and South American	1.15 (0.67-2)	1.21 (0.75-1.96)	1.1 (0.63-1.9)
US-born other Latino	0.48 (0.2-1.15)[c]	0.84 (0.44-1.59)[c]	0.43 (0.2-0.93)[c]
Foreign-born Mexican	0.72 (0.51-1.01)	0.86 (0.64-1.15)	0.51 (0.36-0.72)[d]
Foreign-born Puerto Rican	0.74 (0.41-1.34)	0.86 (0.51-1.46)	0.57 (0.32-1)[c]
Foreign-born Central and South American	0.61 (0.41-0.92)[d]	0.82 (0.59-1.13)	0.7 (0.5-0.99)[d]
Foreign-born other Latino	0.26 (0.1-0.68)[c]	0.6 (0.32-1.12)[c]	0.57 (0.27-1.17)[c]

[a] Control variables not listed include age, sex, education, income to federal poverty level ratio, occupational group, insurance status, daily internet use, and region.

[b] OR: odds ratio.

[c] Too few individuals use the technology in question to draw a statistical conclusion.

[d] Indicates values are significant.

Posthoc Analysis

We chose to analyze Mexican-identified individuals because they are the largest Hispanic ethnic subgroup in the United States [34] and contain enough individuals to do a subgroup analysis of a single Latino subethnic group (N=5310). As age is thought to be a factor in the use of social media and other Web-based communication channels [35], we conducted 2 posthoc analyses of Mexican-identified individuals to see if age was significant in predicting internet use, online HISB, or the use of HIT. We also conducted a third posthoc analysis to see if nativity and insurance status were significant predictors of online HISB and patient portal use among Mexican-identified individuals who used the internet. The supplementary file for posthoc analyses is provided in Multimedia Appendix 1.

Examining Ages 18 to 30 Years

We first examined Mexican-identified individuals to see if individuals aged 18 to 30 years behaved differently than individuals aged 31 years or older (posthoc analysis tables available from author on request). We found that those aged 18 to 30 year had higher odds of using the internet (OR 3.46, 95% CI 2.61-4.59). Among individuals who used the internet, individuals aged 18 to 30 years had lower odds of looking up health information online (OR 0.75, 95% CI 0.58-0.96), filling prescriptions online (OR 0.45, 95% CI 0.27-0.75), and using a chat group to learn about health topics (OR 0.55, 95% CI 0.32-0.96) than individuals aged 31 or older. There was no significant difference between these 2 age groups with regard to emailing a health care provider (OR 0.64, 95% CI 0.4-1.02).

Examining Ages 55 Plus

We then examined Mexicans who used the internet to see if older individuals (aged 55 years or older) were significantly different than other age groups. Individuals who were 55 years or older had lower odds of using the internet (OR 0.13, 95% CI 0.1-0.17) and higher odds of filling prescriptions online (OR 1.91, 95% CI 1.17-3.14). There was no significant difference between individuals who were older than 55 years and individuals aged 18 to 54 years with respect to looking up health information online (OR 0.94, 95% CI 0.71-1.23), scheduling a medical appointment online (OR 1.25, 95% CI 0.77-2.01), emailing a health care provider (OR 1.66, 95% CI 0.97-2.85), or using a chat group to learn about health topics (OR 2.02, 95% CI 0.97-4.18).

Within-Group Nativity Differences

We then examined online HISB among Mexican-identified individuals who used the internet to understand differences in online HISB and patient portal use by nativity. We found that US-born individuals had higher odds (OR 52.9, 95% CI 1.2-1.93) of looking up health information online compared with foreign-born individuals. There was no significant difference between US- and foreign-born Mexicans regarding the use of chat groups to learn about a health topic. US-born Mexican-identified individuals had significantly higher odds of using a computer to fill a prescription (OR 1.58, 95% CI 1.02-2.44) or emailing a provider (OR 1.73, 95% CI 1.13-2.64) than foreign-born Mexican-identified individuals. There was no significant difference between US- and foreign-born Mexicans who used the internet with regard to scheduling a medical appointment online. When we examined the relationship between having insurance and HISB or patient portal use among all Mexican-identified individuals who used the internet, we found that individuals who were insured had higher odds of filling prescriptions online (OR 2.66, 95% CI 1.35-5.25), making medical appointments online (OR 2.08, 95% CI 1.25-3.47), or using email to contact their provider (OR 2.45, 95% CI 1.35-4.46).

Discussion

Principal Findings

Our findings show that although the use of Web-based patient portals is growing [1], there are significant disparities in online HISB and patient portal use in the United States between NH whites and various Latino subgroups after controlling for individual-level differences; however, these disparities vary by subgroup affiliation, nativity, and mode of access. Our findings also imply a gap in access between NH whites and the Mexican subpopulation as a whole, whereas foreign-born groups experience similar disparities across ethnic groups. Although recent reports have shown that internet access among foreign-born Latinos is increasing [17,33], our data show that the Mexican population and foreign-born Latinos in general should be a target for increased access.

Our findings regarding online HISB and patient portals are in line with previous literature showing low internet efficacy among Latinos as compared with whites [8,9,13,36] and showing that Latinos may be less likely to use patient portals [4,37]. However, we expand upon previous research by showing that mode of access, nativity, and subgroup affiliation matter. Not all internet channels are equally likely to reach subpopulations within the Latino population. The need to pay attention to nativity and subgroup or regional affiliation among Latinos is particularly important in light of other studies that show that the internet is becoming the most common source of health information for all ethnic groups [38]. In addition, although emerging research shows that Spanish speakers view Web-based health portals positively [39] and computer-assisted interventions are being deployed more frequently among US Latinos [40], our research implies that the deployment of such interventions online will have uneven penetration among Latinos, particularly foreign-born individuals and Mexican-identified persons. Our findings also support previous research showing that certain portions of the US Mexican-identified population, such as farmworkers, do not seek health information online [36].

Generally, except for US-born Central and South Americans, our respondents are not looking up health information at the same rate as NH whites, although this result is only statistically significant for some groups. The pattern of the direction of the findings for using chat groups is the same for all groups except US-born Puerto Ricans, US-born Central and South Americans, and US- and foreign-born other Latinos. It is only significant for foreign-born Central and South Americans such that they are more likely than NH whites. It may be that this group uses chat groups that allow them to maintain relationships with social networks at home, perhaps allowing people to have contact with others from a similar cultural background. Our data suggest a trend where foreign-born Latinos use patient portals less often than US-born NH whites; however, the difference is only significant for some groups. There are a number of different possible reasons for this. Latinos who are less comfortable with English trust mediated sources less [41]. Previous literature has shown that Latinos may rely on and prefer interpersonal channels [42,43], and 1 study suggested that although some Latinos view patient portals positively, they are less interested in accessing their own medical records online [44], whereas an analysis of focus groups found that Latinos preferred portals that were easy to use and accessible on a mobile phone [43]. Our findings suggest that medical care providers could encourage the use of portals by their patients from these subpopulations by providing handouts, short conversations during appointments, or other aids to encourage use. Alternative forms of patient portals, including portals that are in Spanish, culturally tailored, encourage more personal forms of patient-provider interactions and are easily accessible on mobile phones, which may be a more effective way to reach this target population.

Mexican-Identified Subpopulation

Our posthoc analysis of the Mexican-identified subpopulation shows that nativity also matters within Latinos groups, which is in line with other studies that use Spanish as a measure of acculturation [5,36] and with previous state-level studies showing differences between US- and foreign-born Latinos [9]. US-born individuals within the Mexican subethnicity were generally more likely to use HISB or patient portals than their foreign-born counterparts (although this was not true across the board), which is in line with other studies that show that more acculturated Mexican Americans are willing to use patient portals [44]. Among Mexican-identified persons, individuals who were insured had higher odds of using patient portals more within this group, which may indicate that access to health care is an important factor for patient portal use among Mexican Americans.

Age Effects

Previous studies have shown that individuals who are younger than 35 years are less likely to use patient portals [4], and our posthoc analysis found that younger Mexicans had lower odds of looking up health information online, filling prescriptions online, scheduling a medical appointment online, and using a chat group to learn about health topics, which may be because of the fact that younger individuals have less health problems than older individuals or that individuals with multiple chronic conditions are more likely to use patient portals than those without chronic conditions [45,46]. Although some studies have shown that older Latino adults were less likely to have activated patient portals as compared with NH whites [16], we found that among Mexican-identified individuals, people aged 55 years or older had higher odds of filling prescriptions online, which may be related to their health status. Future research should examine which Web-based apps, programs, and services are being used, especially by foreign-born Latinos and the Mexican-identified population. Such research will allow effective dissemination of health information to foreign-born individuals and potentially reach individuals who are not using other avenues to obtain health information.

Limitations

As this is a secondary data analysis of information, the primary limitation is the use of cross-sectional data, which means there is no causal analysis possible. The data are self-reported; therefore, recall bias and other factors may affect measurement. Furthermore, literacy significantly influences online behavior [47-49], and no measures of online literacy, including health or computer literacy, are assessed in the NHIS. For some groups, too few people reported engaging in HIT to report the findings as statistically significant; consequently, in 1 of our posthoc analyses, we combined US- and foreign-born individuals. Furthermore, as the questions on internet use, HISB, and HIT were asked to all respondents, it was possible that the respondents' answers were inconsistent, where they could answer that they did not use the internet but report that they emailed their provider.

Conclusions

Our findings indicate that the significant gaps in online health-related behaviors between NH whites and Latinos vary by subgroup affiliation, nativity, and mode of access. Although foreign-born Latinos of all ethnic subgroups have lower rates of online HISB, US-born Mexicans also experience lower odds than US-born NH whites. This is important in terms of information dissemination because it suggests that many of the channels that are increasingly being used by NH whites are not effective for communicating with many Latino groups. In addition, our results highlight the need for additional training in internet and patient portal technologies as health information becomes increasingly available online for many of the groups in this study. Without additional training and given that there is such a variation in health behaviors and outcomes, care should be taken when deciding which online channels will be used to reach these populations.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary file for posthoc analyses.

[PDF File (Adobe PDF File), 123KB - jmir_v21i4e10389_app1.pdf]

References

1. Cohen RA, Stussman B. Centers for Disease Control and Prevention. Health Information Technology Use Among Men and Women Aged 18–64: Early Release of Estimates From the National Health Interview Survey, January–June 2009 URL: https://www.cdc.gov/nchs/data/hestat/healthinfo2009/healthinfo2009.pdf [WebCite Cache ID 73MbrcCfp]
2. McCann E. Healthcare IT News. Most providers now using patient portals URL: http://www.healthcareitnews.com/news/most-providers-now-using-patient-portals [accessed 2018-03-13] [WebCite Cache ID 6xsJ8I5ls]
3. 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;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
4. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
5. De Jesus M, Xiao C. Predicting internet use as a source of health information: a “Language Divide” among the Hispanic population in the United States. Policy Internet 2012;4(2):1-11 [FREE Full text] [doi: 10.1515/1944-2866.1178]
6. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
7. Cunningham PJ, Hibbard J, Gibbons CB. Raising low 'patient activation' rates among Hispanic immigrants may equal expanded coverage in reducing access disparities. Health Aff (Millwood) 2011 Oct;30(10):1888-1894 [FREE Full text] [doi: 10.1377/hlthaff.2009.0805] [Medline: 21976331]
8. Brown A, Lopez G, Lopez M. Pew Research Center. 2016. Digital Divide Narrows for Latinos as More Spanish Speakers and Immigrants Go Online URL: http://tinyurl.com/htp3jnv [accessed 2018-08-31] [WebCite Cache ID 6xsIe36Js]
9. Gonzalez M, Sanders-Jackson A, Emory J. Online health information-seeking behavior and confidence in filling out online forms among Latinos: a cross-sectional analysis of the California Health Interview Survey, 2011-2012. J Med Internet Res 2016 Jul 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
10. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-language consumer health information technology interventions: a systematic review. J Med Internet Res 2016 Dec 10;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
11. Christopher Gibbons M. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011;8:1f [FREE Full text] [Medline: 21307989]
12. Perrin A. Pew Research Center. Smartphones help blacks, Hispanics bridge some-but not all-digital gaps with whites URL: http://tinyurl.com/ya8bdrkz [accessed 2018-10-15] [WebCite Cache ID 73CLXer5n]
13. López L, Grant RW. Closing the gap: eliminating health care disparities among Latinos with diabetes using health information technology tools and patient navigators. J Diabetes Sci Technol 2012 Jan;6(1):169-176 [FREE Full text] [doi: 10.1177/193229681200600121] [Medline: 22401336]
14. Peña-Purcell N. Hispanics' use of internet health information: an exploratory study. J Med Libr Assoc 2008 Apr;96(2):101-107 [FREE Full text] [doi: 10.3163/1536-5050.96.2.101] [Medline: 18379664]
15. 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]
16. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
17. 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 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
18. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of electronic portal use among vulnerable patients in a nationwide practice-based research network: From the OCHIN Practice-Based Research Network (PBRN). J Am Board Fam Med 2016 Oct;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
19. Finney Rutten LJ, Hesse BW, Moser RP, Ortiz Martinez AP, Kornfeld J, Vanderpool RC, et al. Socioeconomic and geographic disparities in health information seeking and internet use in Puerto Rico. J Med Internet Res 2012;14(4):e104 [FREE Full text] [doi: 10.2196/jmir.2007] [Medline: 22849971]
20. Lee YJ, Boden-Albala B, Larson E, Wilcox A, Bakken S. Online health information seeking behaviors of Hispanics in New York City: a community-based cross-sectional study. J Med Internet Res 2014;16(7):e176 [FREE Full text] [doi: 10.2196/jmir.3499] [Medline: 25092120]
21. Lee YJ, Boden-Albala B, Quarles L, Wilcox A, Bakken S. Predictors of health information-seeking behaviors in Hispanics. NI 2012 (2012) 2012;2012:243 [FREE Full text] [Medline: 24199094]
22. Selsky C, Luta G, Noone A, Huerta EE, Mandelblatt JS. Internet access and online cancer information seeking among Latino immigrants from safety net clinics. J Health Commun 2013;18(1):58-70 [FREE Full text] [doi: 10.1080/10810730.2012.688248] [Medline: 23066874]

23. Stepler R, Lopez M. Pew Research Center. US Latino Population Growth and Dispersion Has Slowed Since Onset of the Great Recession URL: http://tinyurl.com/jz8yh66 [accessed 2018-10-21] [WebCite Cache ID 6xsHS7h1o]
24. Patten E. Pew Research Center. The Nation's Latino Population is Defined by its Youth Nearly Half of US-Born Latinos Are Younger Than 18 URL: http://www.pewresearch.org/wp-content/uploads/sites/5/2016/04/PH_2016-04-20_LatinoYouth-Final.pdf [accessed 2018-10-16] [WebCite Cache ID 73DxLdtPw]
25. Aponte J. Diabetes-related risk factors across Hispanic subgroups in the Hispanic health and nutritional examination survey (1982-1984). Public Health Nurs 2009;26(1):23-38. [doi: 10.1111/j.1525-1446.2008.00741.x] [Medline: 19154190]
26. Oquendo MA, Lizardi D, Greenwald S, Weissman MM, Mann JJ. Rates of lifetime suicide attempt and rates of lifetime major depression in different ethnic groups in the United States. Acta Psychiatr Scand 2004 Dec;110(6):446-451. [doi: 10.1111/j.1600-0447.2004.00404.x] [Medline: 15521829]
27. Kerner JF, Breen N, Tefft MC, Silsby J. Tobacco use among multi-ethnic Latino populations. Ethn Dis 1998;8(2):167-183. [Medline: 9681283]
28. Fox RS, Mills SD, Roesch SC, Sotres-Alvarez D, Gonzalez P, Bekteshi V, et al. Perceptions of cancer risk/efficacy and cancer-related risk behaviors: results from the HCHS/SOL Sociocultural Ancillary Study. Health Educ Behav 2018 Oct;45(5):790-799. [doi: 10.1177/1090198117744242] [Medline: 29284295]
29. Fox S, Livingston G. Pew Research Center. 2018. Latinos Online URL: http://www.pewhispanic.org/2007/03/14/latinos-online/ [accessed 2018-10-15] [WebCite Cache ID 73DxrtF6x]
30. Center for Disease Control and Prevention. Centers for Disease Control and Prevention. 2015 NHIS Questionnaire - Sample Adult URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2015/english/qadult.pdf [WebCite Cache ID 6xsJq2D72]
31. Pew Research Center. 2013. Second-Generation Americans: A Portrait of the Adult Children of Immigrants URL: http://www.pewsocialtrends.org/2013/02/07/second-generation-americans/ [accessed 2018-03-04] [WebCite Cache ID 6xsK7zKpP]
32. Pew Research Center. 2013. Between Two Worlds: How Young Latinos Come of Age in America URL: http://tinyurl.com/yyrpj4j2 [accessed 2018-03-03] [WebCite Cache ID 6xsKWp8s3]
33. Brown A, Lopez G, Lopez M. Pew Research Center. 2016. Internet use among Hispanics URL: http://www.pewhispanic.org/2016/07/20/1-internet-use-among-hispanics/ [accessed 2018-03-05] [WebCite Cache ID 6xsGyUHsk]
34. Flores A. Pew Research Center. 2017. How the US Hispanic population is changing URL: http://www.pewresearch.org/fact-tank/2017/09/18/how-the-u-s-hispanic-population-is-changing/ [accessed 2018-10-29] [WebCite Cache ID 73XXuuHfK]
35. Pew Research Center. 2017. Social media use by age URL: http://www.pewinternet.org/chart/social-media-use-by-age/ [accessed 2018-10-29] [WebCite Cache ID 73XY4gcsM]
36. Ginossar T. Predictors of online cancer prevention information seeking among patients and caregivers across the digital divide: a cross-sectional, correlational study. JMIR Cancer 2016 Mar 09;2(1):e2 [FREE Full text] [doi: 10.2196/cancer.5108] [Medline: 28410177]
37. Ancker JS, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. Am J Manag Care 2016 Aug;22(8):539-540 [FREE Full text] [Medline: 27541700]
38. Nguyen AB, Robinson J, O'Brien EK, Zhao X. Racial and ethnic differences in tobacco information seeking and information sources: findings from the 2015 Health Information National Trends Survey. J Health Commun 2017 Sep;22(9):743-752. [doi: 10.1080/10810730.2017.1347216] [Medline: 28762887]
39. Ochoa A, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. J Am Med Inform Assoc 2017 Nov 01;24(6):1160-1164. [doi: 10.1093/jamia/ocx040] [Medline: 28460130]
40. Klein CH, Kuhn T, Altamirano M, Lomonaco C. C-SAFE: a computer-delivered sexual health promotion program for Latinas. Health Promot Pract 2017 Dec;18(4):516-525. [doi: 10.1177/1524839917707791] [Medline: 28490203]
41. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. J Health Commun 2010;15(Suppl 3):252-263 [FREE Full text] [doi: 10.1080/10810730.2010.522697] [Medline: 21154097]
42. Elder JP, Ayala GX, Parra-Medina D, Talavera GA. Health communication in the Latino community: issues and approaches. Annu Rev Public Health 2009;30:227-251. [doi: 10.1146/annurev.publhealth.031308.100300] [Medline: 19296776]
43. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": understanding barriers to portal use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
44. Trubitt M, Alozie O, Shokar G, Flores S, Shokar NK. Patterns and correlates of internet use, cell phone use, and attitudes toward patient portals among a predominantly Mexican-American clinic population. Telemed J E Health 2018;24(11):861-869. [doi: 10.1089/tmj.2017.0266] [Medline: 29393769]
45. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
46. Hopman WM, Harrison MB, Coo H, Friedberg E, Buchanan M, VanDenKerkhof EG. Associations between chronic disease, age and physical and mental health status. Chronic Dis Can 2009;29(3):108-116 [FREE Full text] [Medline: 19527569]

47. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
48. Dincer S. A study of the relationship between pupils and parents' computer literacy level and use. Procedia Social Behav Sci 2012;46:484-489 [FREE Full text] [doi: 10.1016/j.sbspro.2012.05.146]
49. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]

Abbreviations

FPL: federal poverty level
HISB: health information–seeking behavior
HIT: health information technology
NH: non-Hispanic
NHIS: National Health Interview Survey
OR: odds ratio

Edited by G Eysenbach; submitted 13.03.18; peer-reviewed by A Nguyen, K Fuji; comments to author 30.07.18; revised version received 31.10.18; accepted 12.12.18; published 16.04.19.

Please cite as:
Gonzalez M, Sanders-Jackson A, Wright T
Web-Based Health Information Technology: Access Among Latinos Varies by Subgroup Affiliation
J Med Internet Res 2019;21(4):e10389
URL: https://www.jmir.org/2019/4/e10389/
doi:10.2196/10389
PMID:30990462

Original Paper

Comparison of Smartphone Ownership, Social Media Use, and Willingness to Use Digital Interventions Between Generation Z and Millennials in the Treatment of Substance Use: Cross-Sectional Questionnaire Study

Brenda L Curtis[1*], PhD; Robert D Ashford[1,2*], MSW; Katherine I Magnuson[3*], BA; Stacy R Ryan-Pettes[3*], PhD

[1]Psychology-Addictions Treatment Research Center, University of Pennsylvania, Philadelphia, PA, United States
[2]Substance Use Disorders Institute, University of the Sciences, Pennsylvania, PA, United States
[3]Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States
[*]all authors contributed equally

Corresponding Author:
Stacy R Ryan-Pettes, PhD
Department of Psychology and Neuroscience
Baylor University
One Bear Place 97334
Waco, TX, 76798-7334
United States
Phone: 1 2547102544
Email: stacy_ryan@baylor.edu

Abstract

Background: Problematic substance use in adolescence and emerging adulthood is a significant public health concern in the United States due to high recurrence of use rates and unmet treatment needs coupled with increased use. Consequently, there is a need for both improved service utilization and availability of recovery supports. Given the ubiquitous use of the internet and social media via smartphones, a viable option is to design digital treatments and recovery support services to include internet and social media platforms.

Objective: Although digital treatments delivered through social media and the internet are a possibility, it is unclear how interventions using these tools should be tailored for groups with problematic substance use. There is limited research comparing consumer trends of use of social media platforms, use of platform features, and vulnerability of exposure to drug cues online. The goal of this study was to compare digital platforms used among adolescents (Generation Zs, age 13-17) and emerging adults (Millennials, age 18-35) attending outpatient substance use treatment and to examine receptiveness toward these platforms in order to support substance use treatment and recovery.

Methods: Generation Zs and Millennials enrolled in outpatient substance use treatment (n=164) completed a survey examining social media use, digital intervention acceptability, frequency of substance exposure, and substance use experiences. Generation Zs (n=53) completed the survey in July 2018. Millennials (n=111) completed the survey in May 2016.

Results: Generation Zs had an average age of 15.66 (SD 1.18) years and primarily identified as male (50.9%). Millennials had an average age of 27.66 (SD 5.12) years and also primarily identified as male (75.7%). Most participants owned a social media account (Millennials: 82.0%, Generation Zs: 94.3%) and used it daily (Millennials: 67.6%, Generation Zs: 79.2%); however, Generation Zs were more likely to use Instagram and Snapchat, whereas Millennials were more likely to use Facebook. Further, Generation Zs were more likely to use the features within social media platforms (eg, instant messaging: Millennials: 55.0%, Generation Zs: 79.2%; watching videos: Millennials: 56.8%, Generation Zs: 81.1%). Many participants observed drug cues on social media (Millennials: 67.5%, Generation Zs: 71.7%). However, fewer observed recovery information on social media (Millennials: 30.6%, Generation Zs: 34.0%). Participants felt that social media (Millennials: 55.0%, Generation Zs: 49.1%), a mobile phone app (Millennials: 36.9%, Generation Zs: 45.3%), texting (Millennials: 28.8%, Generation Zs: 45.3%), or a website (Millennials: 39.6%, Generation Zs: 32.1%) would be useful in delivering recovery support.

Conclusions: Given the high rates of exposure to drug cues on social media, disseminating recovery support within a social media platform may be the ideal just-in-time intervention needed to decrease the rates of recurrent drug use. However, our results suggest that cross-platform solutions capable of transcending generational preferences are necessary and one-size-fits-all digital interventions should be avoided.

(J Med Internet Res 2019;21(4):e13050) doi:10.2196/13050

KEYWORDS

social media, mHealth, substance use treatment, digital health, recovery; social networking sites; substance use disorder

Introduction

Substance use and substance use disorder (SUD) among adolescents and young adults is a major public health concern in the United States. The high rates of substance use coupled with significant unmet treatment needs and alarming rates of recurrence of use are concerning. Recent estimates suggest that 2.0 million adolescents aged 12-17 years and 8.3 million young adults aged 18-25 years used illicit substances in 2017. Among these users, 1.0 million of the adolescent users and 5.2 million of the young adult users were identified as needing SUD treatment. However, only 184,000 of the former and 641,000 of the latter received the required treatment [1]. Strikingly, among those who complete treatment, research suggests that 60%-70% will have a recurrence of use within 90 days after a treatment episode and 85%, within 1 year following treatment [2-4].

These rates of substance use, the unmet need for treatment, and the recurrence of use following treatment suggest that there is a critical need for effective ways to increase service utilization and to make recovery tools available to adolescents and young adults. One way to increase the use of treatment services for SUD among this population is to leverage technologies that are very engaging and already widely used. Given the widespread use of the internet via smartphones and the ubiquitous use of social media, a viable option is to design digital treatments and recovery supports to include internet and social media platforms.

According to new research by the Pew Institute, almost all adolescents and young adults in the United States have access to a smartphone (95% of those aged 13-17 years, 94% of those aged 18-29 years, and 89% of those aged 30-49 years) [5,6]; research suggests that these smartphone owners use the internet extensively. For example, 89% of adolescents aged 13-17 years report using the internet via a mobile device almost constantly or several times a day. Among adults, 89% access the internet daily and 31% access the internet constantly [7]. Noteworthily, visiting social media platforms appears to be important to adolescents and young adults using the internet. Recent results by the Pew Institute [5,8] found that over 92% of adolescents and 88% of adults in the United States use social media platforms. Teens frequently use social media platforms, with 70% using them more than once daily, 38% using them multiple times an hour, and 16% using them near constantly [9]. Among adolescents, the most frequently used social media platforms are YouTube (85%), Instagram (72%), and Snapchat (69%) [5]. In contrast, YouTube and Facebook are used most often among adults (those aged 18-29 years: 91% for YouTube and 81% for Facebook; those aged 30-39 years: 85% for YouTube and 78% for Facebook) [8]. Further, the majority of adult social media users visit these sites very often, with 74% reporting use of Facebook several times a day or at least once a day and 46% reporting use of YouTube several times a day or at least once a day [8].

Designing digital treatments to include the internet and social media platforms is not a novel idea. In fact, in populations of youth with diabetes and obesity, treatment modalities and recovery support through social media were shown to be feasible and effective [10-12]. More specifically, in a sample of 20 adolescents between the ages of 12 and 16 years with type 1 diabetes, those who used a mobile health app with a social networking component showed an improvement in blood glucose monitoring [10]. In addition, preliminary findings in a sample of 13 youth with obesity indicated that social media groups are an acceptable and effective adjunct to obesity treatment [11]. Social media as an intervention has also been used to target sexual health and tobacco cessation. For example, a clustered randomized control trial with 1578 young adults found Facebook an effective medium to disseminate health education [13]. In addition, Facebook was an effective intervention medium for smoking cessation in a sample of 79 young adults [14]. More specifically, this sample of young adults receiving a smoking cessation intervention via Facebook reduced its cigarette consumption by 50% [14].

Although this research demonstrates that the use of the internet and social media to deliver interventions is not new, there are two gaps in the relevant literature. One, our review failed to find reports of formative work that informed the selection of social media platforms and features based on consumer trends. The features included in digital interventions and the selected platforms are likely to vary by population and consumer trends. Popularity of digital platforms and usage trends appear to ebb and flow. For example, until recently, Facebook was the most visited social media site among youth and young adults [5]. Additionally, factors that are related to technology acceptance and social media use are influenced by age [15,16]. Two, there is a dearth of knowledge about the online behavior and preferences for digital treatment and recovery programs in populations enrolled in substance use treatment programs. Although findings of research on adults who use illicit substances are available [17], our review of the literature revealed a lack of knowledge about the use of digital platforms among adolescents attending outpatient substance use treatment. Substance use researchers have begun to explore the use of mobile phones to support recovery from SUDs among adolescents [18]. As researchers move beyond the use of the basic capabilities of mobile phones (eg, texting) and take advantage of more dynamic features of digital and social media platforms [19], it is important to first understand the online behavior and preferences for this type of treatment and recovery support among adolescents. Further, since the development of treatment programs for adolescents tend to trickle down from adult research, it is important to have knowledge of access to mobile phones and the use of digital and social media platforms among adolescents in substance use treatment as compared to other generations in substance use treatment. A better understanding of preferences and use of these platforms across generations of substance users may help with the efforts to tailor technology and digital media platform-based recovery programs. The technology-acceptance model for social media use [15] and evidence showing age as a moderator of mobile health service adoption [16] suggest evident differences in platform use and preferences for key features across generations because younger populations are better able to adapt to novel operating procedures, which increase the perceived ease of use.

Comparing Generation Z and Millennials may be particularly informative. Millennials were coming of age during the proliferation of technology into homes and during the first major introduction of social media into popular culture [20]. Thus, Millennials have adopted technology and social media more than the older generations [21]. On the other hand, Generation Zs were born into a social media culture, positively impacting key variables related to digital and social media acceptance, including perceived ease of use, subjective norm, trust, and reduced risk [15,16]. The purpose of this study is to fill these gaps in the literature by examining and comparing characteristics of various digital platforms used among adolescents (Generation Zs) and young adults (Millennials) attending outpatient substance use treatment and to examine the degree to which these generations are receptive to using these platforms to support substance use treatment and recovery.

Methods

Participants

Adolescents (Generation Zs) and young adults (Millennials) enrolled in outpatient substance use treatment programs in the Southwest and Northeast and regions of the United States, respectively, participated in this study. The requirements to participate in the survey were enrollment in an outpatient program at the time of the survey (all participants), age of 18-35 years (Millennials) or 13-17 years (Generation Zs), no intellectual or developmental disability (all participants), and willingness to provide informed consent or assent to participate. All study procedures were approved by the University of Pennsylvania Human Subjects Review Board and the UT Health San Antonio and Baylor University Institutional Review Boards.

Data for the current study comprise two collated datasets (Generation Zs and Millennials) using an identical survey instrument. The Millennial dataset (n=111) is a subset (those aged between 18 and 35 years only) of a larger sample (n=259) previously reported [17]. The Generation Z dataset has not been previously reported. Following recruitment and data collection, cleaned data were combined into a single data set for analysis.

Procedure

Recruitment of Generation Zs was completed in July 2018, while data from the Millennials were gathered in May 2016. All participants attending outpatient treatment for an SUD completed self-administered, in-person, paper-and-pencil surveys. All participants were invited to participate in this study before, after, or between scheduled group treatment sessions. Participation in this study was voluntary, and it was made clear to all participants that agreeing to be part of this study would not have an impact on their treatment or on any legal proceeding that may have required them to be in treatment. Participants provided individual consent to participate, and where applicable (due to age), participants' parents or legal guardians also provided consent for their child to participate. The survey took 10-15 minutes to complete, and no identifiable information was recorded in order to protect participant privacy.

Measures

The survey used is identical to the survey described by Ashford et al [17] and included questions about technology ownership and use (eg, mobile phone, internet, and social media) and soliciting acceptability and willingness to participate in SUD interventions delivered via digital platforms (Multimedia Appendix 1).

Technology Ownership and Social Media Use

To facilitate comparison across studies in the substance use literature, technology ownership and use were measured using modified questions developed by McClure and colleagues [22] and widely used by the Pew Research Center, a leading authority on trends in mobile phone technology and internet and social media use. We measured social media account ownership and use via novel questions following a structure similar to that developed previously by McClure [22]. All ownership and use questions relied upon self-reported information from participants.

Acceptability and Willingness to Participate in a Digital Intervention

We measured participant willingness to use online platforms for interventions that promote positive recovery outcomes through responses to the following locally developed items: (1) "Do you think social media would be a good place to receive information to help you avoid relapse?" (binary; yes/no), (2) "Would you join an online support group to help you during your recovery?" (binary; yes/no), (3) "Would you join a Facebook support group to help you during your recovery?" (binary; yes/no), (4) "Would you sign up to receive text messages to help you during your recovery?" (binary; yes/no), and (5) "Would you use an app placed on your phone to help your recovery from alcohol or substance use?" (binary; yes/no). We also asked participants to identify the platform they would most like to use in order to access a digital support program to aid during recovery (website, social media, texting, and digital app), and if they would allow their social media accounts to be monitored to help prevent relapse (binary; yes/no).

Frequency of Exposure to Drug Cues and Recovery Information

Participant exposure frequency to drug cues (eg, text, still imagery, or video content related to illicit or licit substances) and recovery cues (eg, text, still imagery, or video content related to recovery and wellness) on social media was measured via responses to the following locally developed items: (1) “How often have you seen drug cues—things that made you want to use drugs on social media?” (Likert scale: 1 [always] to 5 [never]), (2) “How often have you seen recovery information on social media?” (Likert scale: 1 [always] to 5 [never]), and (3) “Have you posted information on social media about being in recovery?” (binary; yes/no).

Substance Use Preferences and Experiences

Participants’ preferences and experiences related to past substance use were collected through a combination of the Alcohol Use Disorder Identification Test - Alcohol Consumption Questions (AUDIT-C) [23], the Drug Abuse Screen Test (DAST-10) [24], and a single self-report question asking participants if a substance was their preferred substance of use (alcohol, opioids, cocaine, amphetamines, marijuana, or other). For this study, we did not use standardized scoring of the AUDIT-C or the DAST-10. Participants were recruited from SUD treatment settings and were presumably already provided an SUD diagnosis. As we were primarily interested in participant substance use preferences and experiences, rather than the diagnosis, we reported individual question responses for each generation and compared response rates for each.

Data Analysis

Data from Millennial participants are a subset of a larger sample reported previously (n=259) [17], although data from the AUDIT-C and DAST-10 were not previously reported. Generation Z responses were entered into a data-monitoring system using double entry. One research assistant entered the data while checking for mismatches and out-of-range values. A different research assistant then entered the same data again. The two entries were compared via a computer that identified mismatches. When mismatches were identified, the person who entered the data checked the original survey to determine the correct value.

All analyses were performed using SPSS V24.0 (IBM Corp, Armonk, NY). Descriptive statistics were calculated for the total sample as well as the Generation Z and Millennial subsamples. Pearson chi-square tests were used to compare technology ownership, use, potential use, and willingness over the two generational groups. No data were excluded from participants meeting the inclusion criteria of this study.

Results

Participants

Demographic information for all participants is shown in Table 1. Overall, participants (n=164) had a mean age of 23.78 (SD 7.06) years, primarily identified as male (67.7%), and were unemployed or current students (75.6%). Most participants self-reported cannabis as a primary substance of use (78.7%), followed by alcohol (34.1%), opioids (23.8%), cocaine (13.4%), and amphetamines (6.7%).

Generation Zs (aged 12-17 years; n=53) had a mean age of 15.66 (SD 1.18) years, primarily identified as male (50.9%), and were unemployed or current students (88.7%). Most Generation Zs self-reported cannabis as a primary substance of use (98.1%), followed by alcohol (32.1%), opioids (11.3%), cocaine (9.4%), and amphetamines (7.5%).

Millennials (aged 18-35 years; n=111) had a mean age of 27.66 (SD 5.12) years, primarily identified as male (75.7%), and were unemployed or current students (69.4%). Most Millennials self-reported cannabis as a primary substance of use (69.4%), followed by alcohol (35.1%), opioids (29.7%), cocaine (15.3%), and amphetamines (6.3%).

Technology Ownership and Use

The majority of both generations owned a mobile phone (Millennial [hereafter referred to as M]: 93.7%, Generation Z [hereafter referred to as Z]: 90.6%) that was identified as a smartphone (M: 71.2%, Z: 90.6%), used it regularly (M: 96.4%, Z: 98.1%), and had either a pay-as-you-go (M: 64.9%, Z: 28.3%) or yearly contract (M: 27.0%, Z: 52.8%). Chi-square testing (Table 2) showed that smartphone ownership and contract type were related to generation, with Generation Zs more likely to own a smartphone than Millennials and Millennials more likely to use pay-as-you-go contracts than Generation Zs. Most participants from both generations had changed their number at least once (M: 70.2%, Z: 75.4%), while a small percentage had changed their number four or more times (M: 11.7%, Z: 15.1%).

The majority of participants from both generations used the internet regularly (M: 86.5%, Z: 96.2%), with most accessing the internet via a mobile phone (M: 72.1%, Z: 50.9%). Chi-square tests found that the way in which the internet was accessed was related to generation, with Millennials more likely to use a mobile phone than Generation Zs, and Generation Zs more likely to use a computer at home or another method of access.

The majority of all participants also regularly used a computer (M: 56.8%, Z: 54.7%), email (M: 72.1%, Z: 54.7%), and text messaging (M: 93.7%, Z: 94.3%). The generation was found to be related to only the regular use of email, but the Millennials were more likely to use email than Generation Zs.

Social Media Ownership and Use

Most participants owned a social media account (M: 82.0%, Z: 94.3%) and used it daily (M: 67.6%, Z: 79.2%) or weekly (M:15.3%, Z: 15.1%). Prominent social media platforms used included Facebook (M: 80.2%, Z: 66.0%), Instagram (M: 61.3%, Z: 83.0%), Twitter (M: 27.0%, Z: 26.4%), Google+ (M: 29.7%, Z: 22.6%), and Snapchat (M: 27.0%, Z: 79.2%). Chi-square tests found that use of Facebook, Instagram, and Snapchat was related to participant generation, with Millennials more likely to use Facebook and Generation Zs more likely to use Instagram and Snapchat.

Table 1. Participant demographic characteristics.

Characteristic	Combined sample (N=164), n (%)	Millennial sample (n=111), n (%)	Generation Z sample (n=53), n (%)
Age (years), mean (SD)	23.78 (7.06)	27.66 (5.12)	15.66 (1.18)
Gender			
Female	53 (32.3)	27 (24.3)	26 (49.1)
Male	111 (67.7)	84 (75.7)	27 (50.9)
Race			
Black	82 (50.0)	70 (63.1)	12 (22.6)
Nonblack	82 (50.0)	41 (36.9)	41 (77.4)
Ethnicity (Latino)	65 (39.6)	24 (21.6)	41 (77.4)
Education level			
Did not complete high school	87 (53.0)	34 (30.6)	53 (100.0)
High school graduate or GED[a]	64 (39.0)	64 (57.7)	0 (0.0)
Two-year degree or more	13 (8.0)	13 (11.7)	0 (0.0)
Employment status			
Employed	40 (24.4)	34 (30.6)	6 (11.3)
Unemployed/studying	124 (75.6)	77 (69.4)	47 (88.7)
Primary substance of use			
Alcohol	56 (34.1)	39 (35.1)	17 (32.1)
Opiates	39 (23.8)	33 (29.7)	6 (11.3)
Cocaine	22 (13.4)	17 (15.3)	5 (9.4)
Amphetamine	11 (6.7)	7 (6.3)	4 (7.5)
Cannabis	129 (78.7)	77 (69.4)	52 (98.1)

[a] GED: general educational development.

Social media was used in a variety of ways by all participants who reported using social media (Table 2). Common activities that social media was used for included instant messaging (M: 55.0%, Z: 79.2%), seeing updates about others (M: 53.2%, Z: 69.8%), watching videos from others (M: 56.8%, Z: 81.1%), reading the news and other information (M: 45.0%, Z: 62.3%), staying in touch with friends and family (M: 64.0%, Z: 86.8%), and finding entertaining content (M: 51.4%, Z: 98.1%). Each of the activities mentioned above was found to be related to the generation, with Generation Zs more likely to use social media to instant message, see updates about others, watch videos from others, find news and information, stay in touch with friends and family, and find entertaining content.

Drug Cues and Recovery Support on Digital Platforms

Most participants had seen drug cues on social media (M: 67.5%, Z: 71.7%), with 22.5% of Millennials seeing drug cues either always or very often and 34.0% of Generation Zs seeing them at the same frequency. Conversely, a higher percentage of both generations had never seen recovery information on social media (M: 30.6%, Z: 34.0%), with 16.2% of Millennials and 26.4% of Generation Zs seeing recovery information either always or very often. Less than one-third of both generations reported having previously posted recovery information on their social media (M: 26.1%, Z: 26.4%). Chi-square tests found that the generation was related to seeing recovery information on social media platforms, with Generation Zs more likely to report seeing recovery information always or very often and Millennials more likely to report seeing recovery information sometimes or rarely.

Although a majority of Millennials believe current social media platforms could be used to prevent recurrence of use (50.5%), a little more than one-third of Generation Zs felt the same way (37.7%), but this value was not statistically significant. When asked what type of platform should be used to deliver recovery support, participants felt that social media (M: 55.0%, Z: 49.1%), a mobile phone app (M: 36.9%, Z: 45.3%), text messages (M: 28.8%, Z: 45.3%), or a website (M: 39.6%, Z: 32.1%) would be useful. Participants' beliefs in the usefulness of texting as a platform were found to be related to generation, with Generation Zs more likely to believe a texting platform would be useful. Less than half the participants from either generation reported a willingness to consent to social media monitoring to support their recovery (M: 36.9%, Z: 24.5%).

Table 2. Technology ownership and usage characteristics by generation.

Ownership and usage characteristics	Millennial sample (n=111), n (%)	Generation Z sample (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
Mobile phone ownership			0.5 (1)	.47
Yes	104 (93.7)	48 (90.6)		
No	7 (6.3)	5 (9.4)		
Smartphone ownership			9.5 (1)	.002[a]
Yes	79 (71.2)	48 (90.6)		
No	32 (28.8)	5 (9.4)		
Mobile phone regular use			0.4 (1)	.55
Yes	107 (96.4)	52 (98.1)		
No	4 (3.6)	1 (1.9)		
Mobile phone contract type			19.4 (2)	<.001[b]
Pay as you go	72 (64.9)	15 (28.3)		
Yearly contract	30 (27.0)	28 (52.8)		
Changed phone number			2.1 (4)	.72
Never	31 (27.9)	13 (24.5)		
One time	33 (27.9)	19 (35.8)		
Two times	16 (14.4)	8 (15.1)		
Three times	18 (16.2)	5 (9.4)		
Four or more times	13 (11.7)	8 (15.1)		
Internet access			13.4 (3)	.01[a]
Via mobile phone	80 (72.1)	27 (50.9)		
Via computer at home	12 (10.8)	18 (34.0)		
Via other method	4 (3.6)	6 (11.3)		
Internet regular use			3.7 (1)	.056
Yes	96 (86.5)	51 (96.2)		
No	15 (13.5)	2 (3.8)		
Computer regular use			0.1 (1)	.81
Yes	63 (56.8)	29 (54.7)		
No	48 (43.2)	24 (45.3)		
Email regular use			4.8 (1)	.03[a]
Yes	80 (72.1)	29 (54.7)		
No	31 (27.9)	24 (45.3)		
Text message regular use			0.0 (1)	.87
Yes	104 (93.7)	50 (94.3)		
No	7 (6.3)	3 (5.7)		
Social media account ownership			4.5 (1)	.03[a]
Yes	91 (82.0)	50 (94.3)		
No	20 (18.0)	3 (5.7)		
Social media use frequency			4.3 (3)	.23
Daily	75 (67.6)	42 (79.2)		
Weekly	17 (15.3)	8 (15.1)		

Ownership and usage characteristics	Millennial sample (n=111), n (%)	Generation Z sample (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
Monthly	1 (0.9)	0 (0.0)		
Do not use regularly	18 (16.2)	3 (5.7)		
Social media used for…				
Share photos or videos			0.0 (1)	.95
Yes	77 (69.4)	37 (69.8)		
No	34 (30.6)	16 (30.2)		
Instant message			9.1 (1)	.003[a]
Yes	61 (55.0)	42 (79.2)		
No	50 (45.0)	11 (20.8)		
Share updates about self			2.3 (1)	.13
Yes	56 (50.5)	20 (37.7)		
No	55 (49.5)	33 (62.3)		
Meet new people			2.1 (1)	.15
Yes	51 (45.9)	18 (34.0)		
No	60 (54.1)	35 (66.0)		
See updates about others			4.1 (1)	.04[a]
Yes	59 (53.2)	37 (69.8)		
No	52 (46.8)	16 (30.2)		
Watch videos others post			9.3 (1)	.002[a]
Yes	63 (56.8)	43 (81.1)		
No	48 (43.2)	10 (18.9)		
News and information			4.3 (1)	.04[a]
Yes	50 (45.0)	33 (62.3)		
No	61 (55.0)	20 (37.7)		
Stay in touch with friends and family			9.1 (1)	.002[a]
Yes	71 (64.0)	46 (86.8)		
No	40 (36.0)	7 (13.2)		
Find funny or entertaining content			35.2 (1)	<.001[b]
Yes	57 (51.4)	52 (98.1)		
No	54 (48.6)	1 (1.9)		
Social media platforms used				
Facebook			3.9 (1)	.049[a]
Yes	89 (80.2)	35 (66.0)		
No	22 (19.8)	18 (34.0)		
Twitter			0.0 (1)	.93
Yes	30 (27.0)	14 (26.4)		
No	81 (73.0)	39 (73.6)		
Google+			0.9 (1)	.34
Yes	33 (29.7)	12 (22.6)		
No	78 (70.3)	41 (77.4)		

Ownership and usage characteristics	Millennial sample (n=111), n (%)	Generation Z sample (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
Instagram			7.5 (1)	.005[a]
Yes	68 (61.3)	44 (83.0)		
No	43 (38.7)	9 (17.0)		
Tumblr			0.4 (1)	.51
Yes	7 (6.3)	2 (3.8)		
No	104 (93.7)	51 (96.2)		
Pinterest			0.1 (1)	.71
Yes	8 (7.2)	3 (5.7)		
No	103 (92.8)	50 (94.3)		
Snapchat			39.7 (1)	<.001[b]
Yes	30 (27.0)	42 (79.2)		
No	81 (73.0)	11 (20.8)		
LinkedIn			3.0 (1)	.09
Yes	6 (5.4)	0 (0.0)		
No	105 (94.6)	53 (100.0)		
Myspace			4.0 (1)	.045
Yes	8 (7.2)	0 (0.0)		
No	103 (92.8)	53 (100.0)		
Seen drug cues on social media			4.5 (2)	.49
Always/very often	25 (22.5)	18 (34.0)		
Sometimes/rarely	50 (45.0)	20 (37.7)		
Never	26 (23.4)	11 (20.8)		
Seen recovery information on social media			18.5 (2)	.002[a]
Always/very often	18 (16.2)	14 (26.4)		
Sometimes/rarely	49 (44.1)	13 (24.5)		
Never	34 (30.6)	18 (34.0)		
Post recovery information			0.0 (1)	.97
Yes	29 (26.1)	14 (26.4)		
No	82 (73.9)	39 (73.6)		
Social media should be used to prevent relapse			2.3 (1)	.13
Yes	56 (50.5)	20 (37.7)		
No	55 (49.5)	33 (62.3)		
Type of platform to deliver relapse-prevention support				
Website			0.9 (1)	.35
Yes	44 (39.6)	17 (32.1)		
No	67 (60.4)	36 (67.9)		
Social media			0.5 (1)	.48
Yes	61 (55.0)	26 (49.1)		
No	50 (45.0)	27 (50.9)		
Texting			4.3 (1)	.04[a]
Yes	32 (28.8)	24 (45.3)		

Ownership and usage characteristics	Millennial sample (n=111), n (%)	Generation Z sample (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
No	79 (71.2)	29 (54.7)		
Mobile phone app			1.04 (1)	.31
Yes	41 (36.9)	24 (45.3)		
No	70 (63.1)	29 (54.7)		
Consent to social media monitoring to support recovery			2.5 (1)	.11
Yes	41 (36.9)	13 (24.5)		
No	70 (63.1)	40 (75.5)		

[a] Significant at *P*=.05.

[b] Significant at *P*<.001.

Substance Use Preferences and Experiences

Most participants reported alcohol intake frequency of monthly or less (M: 67.5%, Z: 81.1%) as well as never having heavy drinking episodes (M: 52.3%, Z: 60.4%). Most participants from either generation had engaged in illicit substance use at some point in their life (M: 55.9%, Z: 64.2%), with less than half reporting polysubstance use (M: 31.5%, Z: 47.2%).

Most participants felt they could stop using substances at any time they wanted (M: 53.2%, Z: 77.4%), and this was related to generation, with Generation Zs being more likely than Millennials to believe they could stop using at any time.

Participants reported a variety of experiences related to their substance use (Table 3). More prevalent experiences for both generations included their family complaining about their substance use (M: 55.9%, Z: 81.1%), guilty feelings (M: 48.6%, Z: 45.3%), engagement in illegal activities (M: 48.6%, Z: 32.1%), or experiencing withdrawal symptoms (M: 41.4%, Z: 32.1%). Family complaining about substance use and engagement in illegal activities were related to generation, with Generation Zs more likely to have experienced complaining family members and Millennials more likely to have engaged in illegal activities.

Table 3. Alcohol and other substance use trends and experiences.

Trends and experiences	Millennial sample (n=111), n (%)	Generation Z (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
Alcohol intake frequency			8.3 (4)	.14
Never	38 (34.2)	22 (41.5)		
Monthly or less	37 (33.3)	21 (39.6)		
2-4 times a month	21 (18.9)	9 (17.0)		
2-3 times a week	10 (9.0)	0 (0.0)		
4 or more times a week	5 (4.5)	1 (1.9)		
Heavy drinking frequency			1.9 (4)	.76
Never	58 (52.3)	32 (60.4)		
Less than monthly	35 (31.5)	16 (30.2)		
Monthly	11 (9.9)	3 (5.7)		
Weekly	5 (4.5)	1 (1.9)		
Daily or almost daily	2 (1.8)	1 (1.9)		
Illicit substance use			1.0 (1)	.31
Yes	62 (55.9)	34 (64.2)		
No	49 (44.1)	19 (35.8)		
Polysubstance use			3.8 (1)	.05
Yes	35 (31.5)	25 (47.2)		
No	76 (68.5)	28 (52.8)		
Can stop using substances at any time[b]			9.9 (1)	.007[a]
Yes	59 (53.2)	41 (77.4)		
No	46 (41.4)	12 (22.6)		
Experiences related to substance use				
Blackouts/flashbacks			0.9 (1)	.34
Yes	33 (29.7)	12 (22.6)		
No	78 (70.3)	41 (77.4)		
Guilty feeling[b]			1.8 (1)	.41
Yes	54 (48.6)	24 (45.3)		
No	54 (48.6)	29 (54.7)		
Family complaints			10.0 (1)	.002[a]
Yes	62 (55.9)	43 (81.1)		
No	49 (44.1)	10 (18.9)		
Neglected family members			1.3 (1)	.25
Yes	44 (39.6)	26 (49.1)		
No	67 (60.4)	27 (50.9)		
Engagement in illegal activity			4.0 (1)	.045[a]
Yes	54 (48.6)	17 (32.1)		
No	57 (51.4)	36 (67.9)		
Withdrawal symptoms			1.3 (1)	.25
Yes	46 (41.4)	17 (32.1)		
No	65 (58.6)	36 (67.9)		

Trends and experiences	Millennial sample (n=111), n (%)	Generation Z (n=53), n (%)	Chi-square test	
			X^2 (*df*)	*P* value
Medical problems			0.5 (1)	.46
Yes	22 (19.8)	8 (15.1)		
No	89 (80.2)	45 (84.9)		

[a] Significant at *P*<.05.

[b] Sample size varies slightly due to missing data.

Discussion

Overview

These data provide an examination of digital platform use and willingness to receive treatment and recovery information via social media among a sample of Generation Zs attending outpatient treatment for an SUD compared to Millennials attending outpatient substance use treatment. Further, these data provide an examination of exposure to drug cues juxtaposing recovery information online and on social media. The results of this study demonstrate an important first step in examining user trends, experiences, and preferences to develop digital interventions that are tailored to demographic trends in use of digital mediums and preferred digital features among a substance-using population. Results of this study in conjunction with the follow-up formative work examining responses to key survey items have the potential to offer prescriptive programming for developers of digital recovery tools.

Principal Results

Unsurprisingly, most Generation Zs and Millennials owned a smartphone, which is consistent with the national rates [5]. Compared to Millennials, Generation Zs were more likely to own a smartphone that was part of a "contract" plan. One explanation for this difference is that individuals under the age of 18 years are more likely to have a mobile phone that is covered as part of a single-family contract [25]. Additionally, results showed that Generation Zs were more likely to have dedicated internet access and Millennials mainly relied on their mobile phone to access the internet. An explanation for this difference is that it is more likely that individuals under the age of 18 years are attending schools that have computers with internet access and are more likely to be living at home with a parent or guardian who is able to provide access to multiple forms of technology [26]. Further, results showed that Millennials were more likely to use email than Generation Zs. One explanation for this difference is that email communication was the primary method for rapid communication as Millennials were coming of age [27], whereas texting and social media features were the primary method for rapid communication among Generation Zs [28].

Social media use dominates both age groups, with most respondents having a social media account that was used daily. However, the preferred social media platform differed between Generation Zs and Millennials, with Generation Zs being more likely to use Instagram and Snapchat and Millennials being more likely to use Facebook. These findings are consistent with current national trends [5] and extend previous findings by demonstrating that generational differences in preferred social media platforms can be generalized to adolescents and young adults in treatment for an SUD. Interestingly, while both generations used social media at high rates, Generation Zs used certain features of social media platforms more than Millennials (eg, instant messaging, seeing updates about others, watching videos from others, finding news and information, and finding entertaining content). This finding suggests that Generation Zs may be more likely to use such features if they were included in a digital intervention.

Implications for Designing Digital Interventions

These findings provide important information for developers interested in designing digital treatment and recovery programs for adolescents and young adults with problematic patterns of substance use. Findings that a large percentage of Generation Zs and Millennials in substance use treatment use social media suggests that the use of social media platforms to deliver treatment and recovery-related information would allow practitioners to meet these populations where they are. The most common digital platforms used for disseminating treatment and recovery information are mobile phone apps [19,29]. Results of this study suggest that general dissemination of information via apps may not reach the intended audience. Millennials are not likely to find mobile apps helpful.

When asked what platform would be most helpful to receive recovery support, Generation Zs and Millennials rated social media above other platforms: 49% of Generation Zs and 55% of Millennials thought it would be helpful to receive relapse-prevention support via social media. This response rate is encouraging, considering that the remaining participants likely include a mix of at least four types of people: those who do not need relapse prevention by any method of delivery; those not interested in relapse prevention, in general; those who truly would not find recovery support through social media helpful; and those with privacy concerns. In support of our belief that participants may have privacy concerns related to recovery support delivered on social media, only 25.5% of Generation Zs and 36.9% of Millennials reported they would consent to social media monitoring to support their recovery. Our survey did not provide participants an opportunity to describe reasons for dissenting on either of these questions. Follow-up focus group data are needed to better understand Generation Zs' and Millennials' perceptions, acceptability, risk, and benefit. It is possible that an interactive dialogue could result in more participants agreeing on the helpfulness of recovery information delivered on social media and to allow their account to be monitored.

Results also showed there is unfavorable variance in the rates of exposure to drug cues and recovery-related information among both generations. The gap between exposure to drug cues and recovery-related information presents a unique opportunity to use social media platforms as an ideal location for digital interventions—both primary (eg, dissemination of treatment and recovery-related information) and tertiary (eg, reducing drug-related information) interventions. In total, 70%-75% of adolescents and young adults are on social media platforms multiple times a day, and 16% reported being on these platforms "near constantly" throughout the day [9]. For individuals with an SUD, exposure to drug cues online could serve as a trigger for continued or recurrent use. Immediate access to recovery supports online and within a social media platform may be the ideal just-in-time intervention necessary to decrease rates of recurrence of use.

Importantly, the high use of technology and digital platforms, along with the differences in the type and preference of specific platforms and activities, underscores the importance of program developers in conducting formative research to determine the preferred "location" at which the targeted population would like to receive interventions. For example, if the target audience is adolescents, the results of this study suggest that leveraging existing platforms such as Snapchat or Instagram may be ideal. However, if the target audience is young adults, leveraging Facebook may be more ideal. Further, the extant literature suggests that if the target audience cuts across Generation Zs and Millennials, leveraging YouTube may be ideal [5].

Although we were unable to find prior formative work on the characteristics and pattern of digital media use among adolescents in substance use treatment, results of this study and prior work with adults in treatment for substance use suggest that the use of a digital platform varies by age group [17,22]. Thus, we further recommend that program developers design tools and interventions that are multi- and cross-platform as a default and carefully consider using individual platforms only when there is a specific targeted audience in mind.

Development of digital interventions should be informed by trends in "user features." Differential use of user features will allow for an informed design and dissemination of intervention programs that are more likely to be adopted by the target audience. For example, if the target audience is adolescents, the data suggest that a texting platform solution would be preferred. Similarly, results suggest that using instant messaging functions may be more impactful for targeting adolescents, while it may be less beneficial for targeting Millennials. Further, Generation Z may respond to recovery messages formatted in the way of news as well as exaggerated messages with cartoons that use humor to teach important skills. Additionally, developers should consider delivering therapeutic content and skills that have been produced by youth by using camera phones and that are capable of evoking relatable and realistic presentation of images. Apart from sharing photos or videos and staying in touch with friends and family, Millennials did not overwhelmingly report the use of any one social media feature. This suggest that although Millennials are online and most use social media like Facebook, they may be most interested in social media-based recovery tools that are less dynamic and include more static content like sharing videos and photos and posting messages with updates. This interpretation is consistent with research showing that older generations are more likely to adopt mobile health services that do not include a lot of operating procedures [16].

Results offer other useful information related to intervention content for digital-based intervention developers. For example, while Generation Zs reported higher rates of family complaints about their substance use as a negative experience, Millennials reported higher rates of criminal activity as negative experiences related to their substance use. Interventions should seek to incorporate this information, perhaps, including more family-oriented content for Generation Zs while incorporating decisional balance activities that help adults explore the pros and cons of substance use for Millennials. Generation Zs were also more likely to believe that they could stop using substances at any time. This may be explained by a shorter length of substance use overall, with fewer (if any) unsuccessful cessation attempts. Digital interventions designed for this generation may also benefit from motivational interviewing techniques [30], as those with an SUD are not more likely to have the ability to quit at will as a factor of age [31].

Limitations

Several limitations of this study should be noted, as they may impact the interpretation and generalizability of the findings. There was limited overlap in data-collection periods between the two generations (Generation Z in 2018 and Millennials in 2016), and it is possible that there were changes in the popularity of certain platforms or feature use in this time frame. However, it is important to note that our results are consistent with current national trends [5-8] and provide a statistical test of observed generational differences. Additionally, the different locations where the data were collected (Generation Z data collected in the Southwestern United States and Millennial data collected in the Northeastern United States) may have captured regional (geographic) differences in addition to or instead of a difference between the generational cohorts. However, while research has suggested that geographic differences in social media use and technology ownership exist between rural, urban, and suburban areas, they have significantly declined in the last decade [21]. Finally, an overwhelming majority of participants reported using marijuana. Given the national trends in legalization of marijuana, a study with the power to conduct analyses by the type of illicit drug may provide additional information. It is possible that report of online exposure to drug information and openness to social media-based interventions are related to social attitudes about particular substances.

Future Directions

In addition to the recommendations for researchers developing digital interventions as outlined above, these data point to the next steps in generating key formative research needed to develop dynamic digital interventions capable of delivering just-in-time treatment and recovery supports. With emerging data showing that social media language is related to diagnosis [32] and emerging machine learning techniques capable of predicting items such as county-level binge drinking rates [33], the design of a digital intervention will soon capitalize on the ability to deliver just-in-time interventions by monitoring language used by users of social media. In this study, a low percentage of Generation Zs and Millennials agreed to have their social media accounts monitored for delivering a treatment program or preventing a recurrence of use. Additional research is needed in this area, as research on older adults with problematic substance use has shown that monitoring and tracking digital behavior and activity are acceptable if respondents could control the features of the device or the service monitoring their actions [34,35]. It is likely that with follow-up questions on this topic, Generation Zs and Millennials would reveal similar results and, perhaps, guidance for how to include such a feature in a digital intervention. Future research should also identify the specific factors influencing the decision to allow monitoring, as it is possible that other reasons exist that can be mitigated through purposeful design and enhanced privacy features.

Conclusions

To our knowledge, this study is the first study to compare two generational groups of clients attending treatment for an SUD in terms of the characteristics of technology, internet, and social media use as well as social media exposure to both drug cues and recovery information. Although mobile phone ownership (including smartphones) and the use of internet and social media platforms are high among both generations, the frequency of use and preference for user functions were different. Developers and practitioners may use findings from technology ownership and research to better inform digital interventions to improve adoption and efficacy and reduce attrition. Although the contrasts between generational cohorts suggest that specific features may be more impactful for a certain generation, such as texting platforms for Generation Zs, the overall future interventions would be well positioned to create cross-platform solutions that can be used across digital platforms, including texting, social media, websites, and mobile phone apps. One-size-fits-all digital interventions should be avoided, as cross-platform solutions are likely able to transcend generational preferences and increase adoption. Importantly, data from this study provide support for the need for digital interventions. The significant differences in the exposure to drug cues and recovery-related information on social media present as an immediate opportunity for intervention developers to immediately improve SUD treatment and recovery outcomes for both Generation Zs and Millennials.

Acknowledgments

We would like to thank the individuals and families that participated in this study as well as our dedicated research assistants Emily Adams, Casey Hamilton, and Dan Summers.

Outcomes reported in this publication were supported by the Delivery Systems Reform Incentive Payment Program (085144601.2.101); the National Center for Advancing Translational Sciences; National Institutes of Health (through Grant KL2 TR002646); the National Institute on Drug Abuse (R01 DA039457); and, in part, funds from the Vice Provost for Research at Baylor University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Social media usage survey.

[PDF File (Adobe PDF File), 241KB - jmir_v21i4e13050_app1.pdf]

References

1. Key substance use and mental health indicators in the United States: Results from the National Survey on Drug Use and Health (HHS Publication No. 2017. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.htm [accessed 2019-04-02] [WebCite Cache ID 77I3P94Ii]
2. Cornelius J, Maisto S, Pollock N, Martin C, Salloum I, Lynch K, et al. Rapid relapse generally follows treatment for substance use disorders among adolescents. Addictive Behaviors 2003 Mar;28(2):381-386 [FREE Full text] [doi: 10.1016/S0306-4603(01)00247-7]
3. Sinha R. New findings on biological factors predicting addiction relapse vulnerability. Curr Psychiatry Rep 2011 Oct;13(5):398-405 [FREE Full text] [doi: 10.1007/s11920-011-0224-0] [Medline: 21792580]
4. Barrett H, Slesnick N, Brody JL, Turner CW, Peterson TR. Treatment outcomes for adolescent substance abuse at 4- and 7-month assessments. Journal of Consulting and Clinical Psychology 2001;69(5):802-813. [doi: 10.1037//0022-006X.69.5.802]
5. Anderson M, Jiang J. Pew Internet. 2018. Teens, social media & technology URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-04-02] [WebCite Cache ID 73R2m3pmH]
6. Pew Research Center: Internet & Technology. 2017. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-02] [WebCite Cache ID 72HdjZi3o]

7. Perrin A, Jiang J. Pew Research Center: Internet & Technology. 2018. About a quarter of US adults say they are 'almost constantly' online URL: http://www.pewresearch.org/fact-tank/2018/03/14/about-a-quarter-of-americans-report-going-online-almost-constantly/ [accessed 2018-10-25] [WebCite Cache ID 73R354Qq0]
8. Smith A, Anderson M. Pew Research Center: Internet & Technology. 2018. Social media use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2018-10-25] [WebCite Cache ID 73R3G2qRK]
9. Rideout V, Robb M. Social media, social life: Teens reveal their experiences. San Francisco, CA: Common Sense Media; 2018. URL: https://www.commonsensemedia.org/sites/default/files/uploads/research/2018_cs_socialmediasociallife_fullreport-final-release_2_lowres.pdf [accessed 2019-04-02] [WebCite Cache ID 77I4M7fC3]
10. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
11. Prout Parks E, Moore RH, Li Z, Bishop-Gilyard CT, Garrett AR, Hill DL, et al. Assessing the feasibility of a social media to promote weight management engagement in adolescents with severe obesity: pilot study. JMIR Res Protoc 2018 Mar 19;7(3):e52 [FREE Full text] [doi: 10.2196/resprot.8229] [Medline: 29555623]
12. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
13. Bull S, Levine D, Black S, Schmiege S, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
14. Ramo DE, 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]
15. Rauniar R, Rawski G, Yang J, Johnson B. Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. Journal of Ent Info Management 2014 Feb 04;27(1):6-30. [doi: 10.1108/JEIM-04-2012-0011]
16. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. International Journal of Information Management 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
17. Ashford RD, Lynch K, Curtis B. Technology and social media use among patients enrolled in outpatient addiction treatment programs: cross-sectional survey study. J Med Internet Res 2018 Mar 06;20(3):e84 [FREE Full text] [doi: 10.2196/jmir.9172] [Medline: 29510968]
18. Gonzales R, Hernandez M, Murphy D, Ang A. Youth recovery outcomes at 6 and 9 months following participation in a mobile texting recovery support aftercare pilot study. Am J Addict 2016 Jan;25(1):62-68 [FREE Full text] [doi: 10.1111/ajad.12322] [Medline: 26689171]
19. Bergman B, Claire Greene M, Hoeppner B, Kelly J. Expanding the reach of alcohol and other drug services: Prevalence and correlates of US adult engagement with online technology to address substance problems. Addict Behav 2018 Dec;87:74-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.06.018] [Medline: 29960132]
20. Dimock M. Pew Research Center. 2018. Defining generations: where millennials end and post-millennials begin URL: http://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/ [accessed 2019-02-05] [WebCite Cache ID 75xssnjuv]
21. Pew Research Center: Internet & Technology. 2018. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-11-07] [WebCite Cache ID 73l08AcDF]
22. McClure EA, Acquavita SP, Harding E, Stitzer ML. Utilization of communication technology by patients enrolled in substance abuse treatment. Drug Alcohol Depend 2013 Apr 01;129(1-2):145-150 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.10.003] [Medline: 23107600]
23. Bush K. The AUDIT Alcohol Consumption Questions (AUDIT-C): An effective brief screening test for problem drinking. Arch Intern Med 1998 Sep 14;158(16):1789 [FREE Full text] [doi: 10.1001/archinte.158.16.1789]
24. Skinner H. The drug abuse screening test. Addictive Behaviors 1982 Jan;7(4):363-371 [FREE Full text] [doi: 10.1016/0306-4603(82)90005-3]
25. Pew Internet. 2010. Teens and Mobile Phones URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2010/PIP-Teens-and-Mobile-2010-with-topline.pdf [accessed 2019-03-31] [WebCite Cache ID 77I45prSe]
26. Turner A. Generation Z: technology and social interest. The Journal of Individual Psychology 2015;71(2):103-113. [doi: 10.1353/jip.2015.0021]
27. Pew Research Center. Millennials: A portrait of generation next URL: http://www.pewresearch.org/wp-content/uploads/sites/3/2010/10/millennials-confident-connected-open-to-change.pdf [accessed 2019-02-05] [WebCite Cache ID 75y0C76q9]
28. Lenhart A, Smith A, Anderson M, Duggan M, Perrin A. Pew Research Center: Internet & Technology. 2015. Teens, technology and friendships URL: http://www.pewinternet.org/2015/08/06/teens-technology-and-friendships/ [accessed 2019-02-05] [WebCite Cache ID 75y0MILt3]
29. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [Medline: 30139724]

30. Miller W. Motivational interviewing in service to health promotion. Am J Health Promot 2004;18:A1. [doi: 10.4278/0890-1171-18.4.TAHP-1]
31. Hser Y, Anglin M. Addiction Treatment and Recovery Careers. In: Kelly JF, White WL, editors. Addiction Recovery Management: Theory, Research and Practice. Totowa, NJ: Humana Press; 2011:978-971.
32. Choudhury DM, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Jul 08 Presented at: Proceedings of the 7th International AAAI Conference Weblogs Social Media; July 2013; Boston, MA p. 8-10.
33. Curtis B, Giorgi S, Buffone AEK, Ungar LH, Ashford RD, Hemmons J, et al. Can Twitter be used to predict county excessive alcohol consumption rates? PLoS One 2018;13(4):e0194290 [FREE Full text] [doi: 10.1371/journal.pone.0194290] [Medline: 29617408]
34. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
35. Beninger K, Fry A, Jago N, Lepps H, Nass L, Silvester H. Nat Cen Social Research. 2014. Research using social media; users' views URL: https://www.researchgate.net/publication/261551701_Research_using_Social_Media_Users'_Views [accessed 2019-03-31] [WebCite Cache ID 77I2hnOFV]

Abbreviations

AUDIT-C: Alcohol Use Disorder Identification Test - Alcohol Consumption Questions
DAST-10: Drug Abuse Screen Test
SUD: substance use disorder

Edited by G Eysenbach; submitted 11.12.18; peer-reviewed by J Priebe, J Bian, M Gomis, A Davoudi, A Brown, I Cop; comments to author 07.01.19; revised version received 11.02.19; accepted 17.02.19; published 17.04.19.

Please cite as:
Curtis BL, Ashford RD, Magnuson KI, Ryan-Pettes SR
Comparison of Smartphone Ownership, Social Media Use, and Willingness to Use Digital Interventions Between Generation Z and Millennials in the Treatment of Substance Use: Cross-Sectional Questionnaire Study
J Med Internet Res 2019;21(4):e13050
URL: http://www.jmir.org/2019/4/e13050/
doi:10.2196/13050
PMID:30994464

Original Paper

Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations

Bethany Percha[1], PhD; Edward B Baskerville[1], PhD; Matthew Johnson[1], MS; Joel T Dudley[1], PhD; Noah Zimmerman[1], PhD

Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Bethany Percha, PhD
Icahn School of Medicine at Mount Sinai
770 Lexington Avenue
15th Floor
New York, NY, 10065
United States
Phone: 1 2127317072 ext 27072
Email: bethany.percha@mssm.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/9/e16179/

Abstract

Background: Recent advances in molecular biology, sensors, and digital medicine have led to an explosion of products and services for high-resolution monitoring of individual health. The N-of-1 study has emerged as an important methodological tool for harnessing these new data sources, enabling researchers to compare the effectiveness of health interventions at the level of a single individual.

Objective: N-of-1 studies are susceptible to several design flaws. We developed a model that generates realistic data for N-of-1 studies to enable researchers to optimize study designs in advance.

Methods: Our stochastic time-series model simulates an N-of-1 study, incorporating all study-relevant effects, such as carryover and wash-in effects, as well as various sources of noise. The model can be used to produce realistic simulated data for a near-infinite number of N-of-1 study designs, treatment profiles, and patient characteristics.

Results: Using simulation, we demonstrate how the number of treatment blocks, ordering of treatments within blocks, duration of each treatment, and sampling frequency affect our ability to detect true differences in treatment efficacy. We provide a set of recommendations for study designs on the basis of treatment, outcomes, and instrument parameters, and make our simulation software publicly available for use by the precision medicine community.

Conclusions: Simulation can facilitate rapid optimization of N-of-1 study designs and increase the likelihood of study success while minimizing participant burden.

(J Med Internet Res 2019;21(4):e12641) doi:10.2196/12641

KEYWORDS

n-of-1 studies; computer simulation; patient-specific modeling; precision medicine; cross-over studies; inter-individual biological variation; individual differences

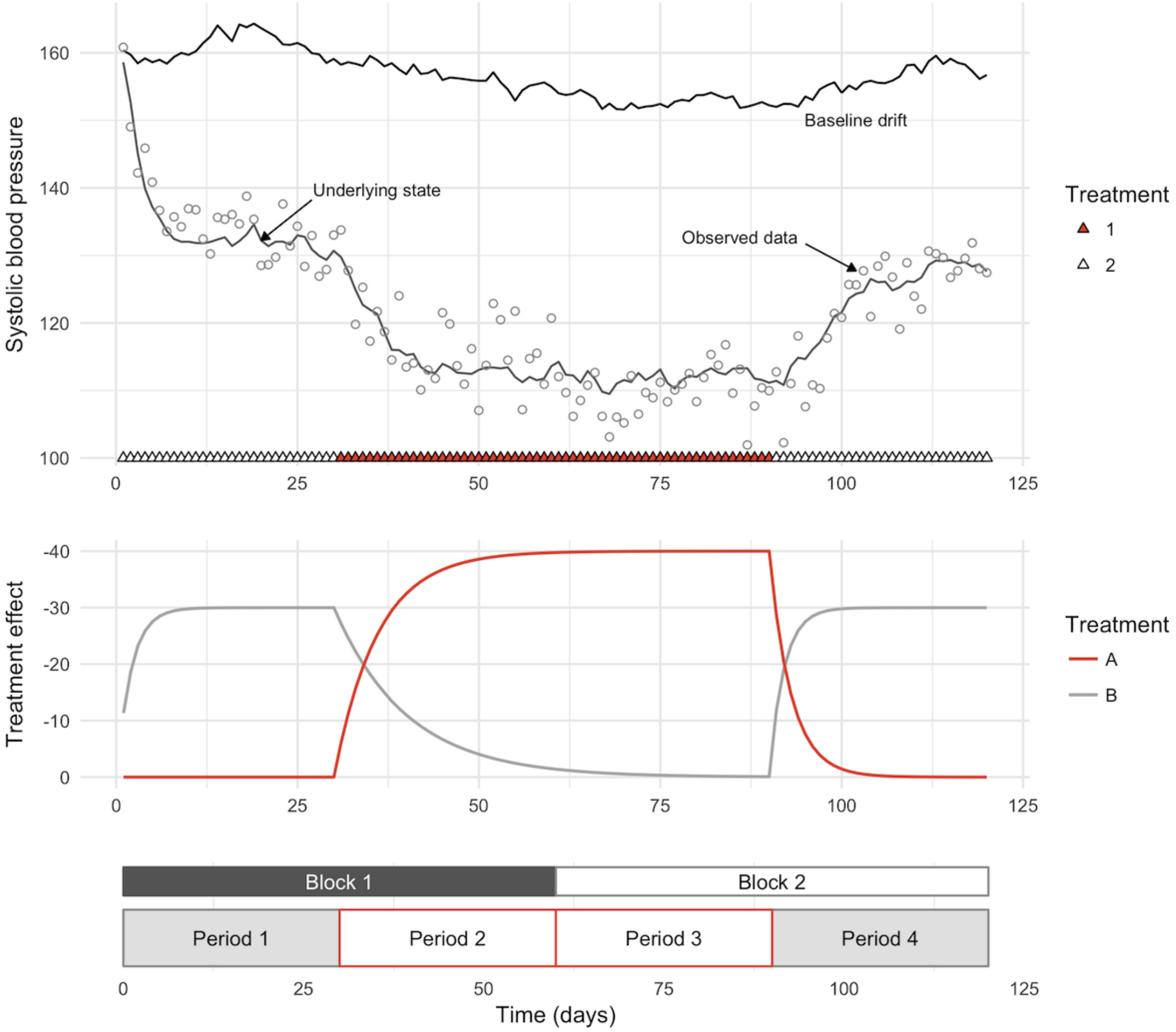
Introduction

The Promise of N-of-1 Studies

N-of-1 studies have shown great promise as a tool for investigating the effects of drugs, supplements, behavioral changes, and other health interventions on individual patients [1-7]. An N-of-1 study (Figure 1) is a multiple-crossover comparative effectiveness study of a single patient. Competing treatments are administered in blocks, within which treatment order is randomized or counterbalanced [6]. The outcome of interest is compared across different treatment periods to find the treatment with the greatest efficacy for that specific patient.

N-of-1 studies inform the care of individual patients while simultaneously generating evidence that can be combined with other N-of-1 studies to yield population-level analyses [8-10]. These studies will likely play a key role in precision medicine, with its focus on narrowly defined patient cohorts, rare conditions, and complex comorbidities [5].

Figure 1. Example of an N-of-1 study comparing two blood pressure medications. An N-of-1 study consists of a set of N blocks, each of which contains J different treatment periods. The order of the treatment periods within each block is usually randomized. Parameters: X0=160, E1=-40, E2=-30, tau1=6.0, gamma1=3.0, tau2=2.0, gamma2=10.0, alpha=0.5, *P*=30, N=2, J=2, sigma_b=0.9, sigma_p=1.0, sigma_0=4.0. In this example, one sample was taken per day.



Challenges to N-of-1 Studies

However, the design and analysis of N-of-1 studies present several methodological challenges. Although the Agency for Healthcare Research and Quality has recently released a set of statistical guidelines for N-of-1 studies [6,11], drawing attention to potential treatment effect confounders like underlying time trends, carryover effects, and autocorrelated measurements, there is currently no universal methodological or statistical framework for the design and analysis of N-of-1 trials. Treatments are often compared graphically or ad hoc measures of efficacy are used that differ from study to study; a review of N-of-1 trials published between 1985 and 2010 found that only 49% used any statistical measure to compare treatments [2]. As a result, it is difficult to compare findings from different studies or understand how specific analytic choices influence study results.

N-of-1 studies must also overcome daunting practical and logistical challenges. For example, although researchers might like to administer treatments over dozens of blocks to increase statistical power, such designs are burdensome to the patient and increase the likelihood of attrition. It is also difficult to convince individuals to revisit earlier treatments, especially if these are perceived as less effective [1,6]. Practically speaking, this means the number of treatment blocks in an N-of-1 study is limited, as is the total duration of the study. Although a statistician might prefer more shorter blocks relative to few longer blocks (since the number of samples in a traditional N-of-1 analysis is linear in the number of blocks), rapid switching among treatments may obscure true differences in efficacy because of carryover effects from earlier treatments. Many treatments, such as antidepressants, also take time to display their full effects. Decisions about the length and arrangement of treatment periods can have a profound effect on statistical effect estimates in N-of-1 studies.

Simulating N-of-1 Studies

Simulation has played a crucial role in clinical trial design, increasing the efficiency and cost-effectiveness of clinical trials, especially in the pharmaceutical industry [12]. Inspired by this, we have developed a stochastic time-series simulation model for N-of-1 studies that incorporates all study-relevant effects, such as carryover and wash-in effects. The model can be used to produce realistic simulated data for a near-infinite number of N-of-1 study designs, treatment profiles, and patient characteristics. The model also incorporates noise parameters like baseline drift, short-term fluctuations (process noise), and measurement error to provide realistic sources of variation that can obscure treatment effects in real-patient settings. Using simulation, we can cheaply and easily investigate how design parameters like sampling frequency, number, and location of samples within blocks, treatment order within blocks, treatment period duration, and total number of blocks impact statistical estimates of treatment effects.

In this paper, we use the model to analyze two N-of-1 case studies, showing how simulation can both optimize study designs and assist researchers in deciding on an appropriate analysis protocol. We then use the model to produce a set of design recommendations for N-of-1 studies on the basis of parameters related to the study outcome, instrument used to measure the outcome, and treatment(s) themselves. We provide our simulation software as a supplement to the paper.

Methods

Stochastic Time-Series Model

Assume that there are J total treatments in an N-of-1 study. Let $B(t)$ denote the patient's true baseline at time t. Let $X_j(t)$ denote the effect of treatment j (j=1, …, J) at time t so that the total treatment effect at time t is $X= \Sigma_j X_j(t)$. Let $T_j(t)$ be 1 if treatment j is in process at time t and 0 otherwise (see Figure 1). Let $Z(t)$ denote the patient's true outcome state at time t, and let $Y(t)$ denote the patient's observed outcome at time t.

The underlying effect driver for each treatment is described as an ordinary differential equation:

$$dX_j = [((E_j - X_j) / \tau_j) T_j(t) - (X_j / \gamma_j) (1 - T_j(t))] dt$$

Here each $X_j(t)$ is an exponential decay toward a target value that changes over time—either E_j or 0, depending on $T_j(t)$—with time constant τ_j during run-in (decay toward E_j) and γ_j during wash-out (decay toward 0).

Baseline drift is simulated as a discretized Wiener process, where normal noise with variance $\sigma_b^2 \Delta t$ is applied every Δt:

$$B(t + \Delta t) = B(t) + \Delta B(t)$$

where

$$\Delta B(t) \sim \text{Normal}(0, \sigma_b^2 \Delta t)$$

The outcome variable $Z(t)$ is also a discrete-time stochastic process,

$$Z(t + \Delta t) = Z(t) + \Delta Z_{det}(t) + \Delta Z_{stoch}(t)$$

where $\Delta Z_{det}(t)$ is a deterministic exponential decay toward the target $X_j(t)+B(t)$:

$$\Delta Z_{det}(t) = Q(t) + [Z(t) - Q(t)] \exp(-\Delta t/\propto)$$

$$Q(t) = B(t) + \Sigma_j X_j(t)$$

with time constant $\propto$ and

$$\Delta Z_{stoch}(t) \sim \text{Normal}(0, \sigma_p^2 \Delta t)$$

The observed outcome differs from the true outcome only through the addition of normally distributed observation noise:

$$Y(t) \sim \text{Normal}(Z(t), \sigma_o)$$

All of the model parameters are summarized in Table 1. Transformations of $Y(t)$ can be used to model different types of outcome parameters, such as scores, counts, and binary outcomes (Table 2).

Table 1. The parameters underlying data generation for an N-of-1 study. The parameters are divided into study design parameters (D), treatment-related parameters (T), measurement parameters (M), and outcome-related parameters (O).

Parameter	Type	Description
$\{t_1,\ldots,t_n\}$	D	Sampling times
N	D	Number of blocks (each with J periods in random order)
J	D	Number of treatment periods per block
P	D	Treatment period length
$E_1,\ldots,E_J$	T	Effect sizes for treatments 1 through J
$\tau_1,\ldots, \tau_J$	T	Run-in time constants for treatments 1 through J
$\gamma_1,\ldots, \gamma_J$	T	Wash-out time constants for treatments 1 through J
$\propto$	O	Sensitivity to treatment effect
σ_b^2	O	Variance of baseline drift process
σ_p^2	O	Variance of process noise
σ_o^2	M	Variance of observation noise

Table 2. Suggested transformations of *Y* for simulating discrete outcomes.

Outcome type	Range of outcome	Distribution of Y	Transformation
Numeric	Real numbers	—[a]	Identity
Score	[0,…,M]	—	Identity (round, truncate)
Count	[0,…,infinity)	Poisson(λ)	$\lambda = \exp(Y)$
Proportion	[0,…,M]	Binomial(M, p)	$P=1/(1 + \exp(-Y))$
Binary	{0, 1}	Bernoulli(p)	$P=1/(1 + \exp(-Y))$

[a] Not applicable.

Hypertension Case Study

A sample data set and all parameter values for the hypertension case study can be found in Figure 1. The study involves 2 different blood pressure medications, one of which reduces systolic blood pressure by 10 more points than the other in the long run. The more effective medication, treatment 1, takes longer to reach its full effect (τ_1=6.0, τ_2=2.0) and less time to wash out (γ_1=3.0, γ_2=10.0). The sampling rate is 1 sample/day, which we chose to model blood pressure that is monitored using a cuff.

We chose a statistical model for this study that incorporated fixed effects for both block ID and treatment, on the basis of the recommendations provided by the Agency for Healthcare Research and Quality (AHRQ) and others [6,11]:

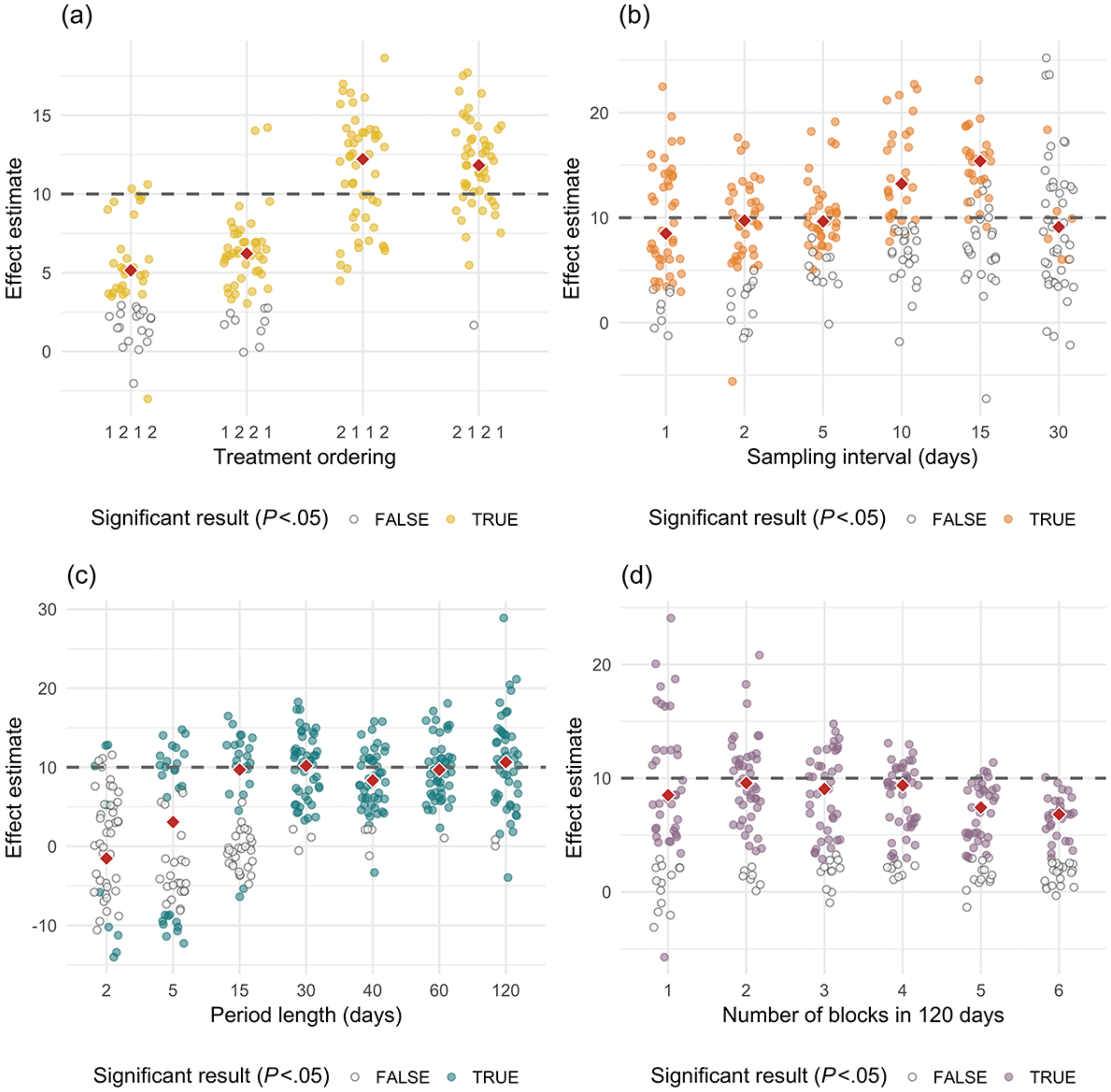
$$y=\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_N x_N$$

where x_1 is 1 if treatment 2 is in progress at the time of the sample, and 0 otherwise, and x_n is 1 if block *n* is in progress, and 0 otherwise. Note that there are only $n-1$ indicator variables for blocks; block 1 is used as the reference block. We experimented with other models but found that although modeling choices could affect power, effect size estimates did not change much among models. Our software provides the ability to choose from among several different models.

To create Figure 2, we repeated the data generation and analysis process, varying the following parameters and keeping the rest constant:

1. Treatment period orderings were varied among 1 2 1 2, 1 2 2 1, 2 1 1 2, and 2 1 2 1.
2. Sampling frequency was varied from 1 sample per day to 1 sample per treatment period, holding the treatment period ordering fixed at 2 1 2 1.
3. Upon holding sampling frequency constant at 1 sample per day, period length was varied from 2 to 120 days.
4. Study length was held constant at 120 days, and the number of blocks was varied from 1 to 6.

Figure 2. Variation in effect estimates for the hypertension study by study design parameters, including (a) treatment period ordering, (b) sampling frequency, (c) treatment period length, and (d) number of blocks for a fixed study length. The true effect size is 10, illustrated by the dashed lines in the figures. The red diamonds correspond to the median effect size for the statistically significant results within each group. Power estimates were obtained by calculating the ratio of the number of colored dots to the number of total dots. There are 50 trials shown for each parameter setting.

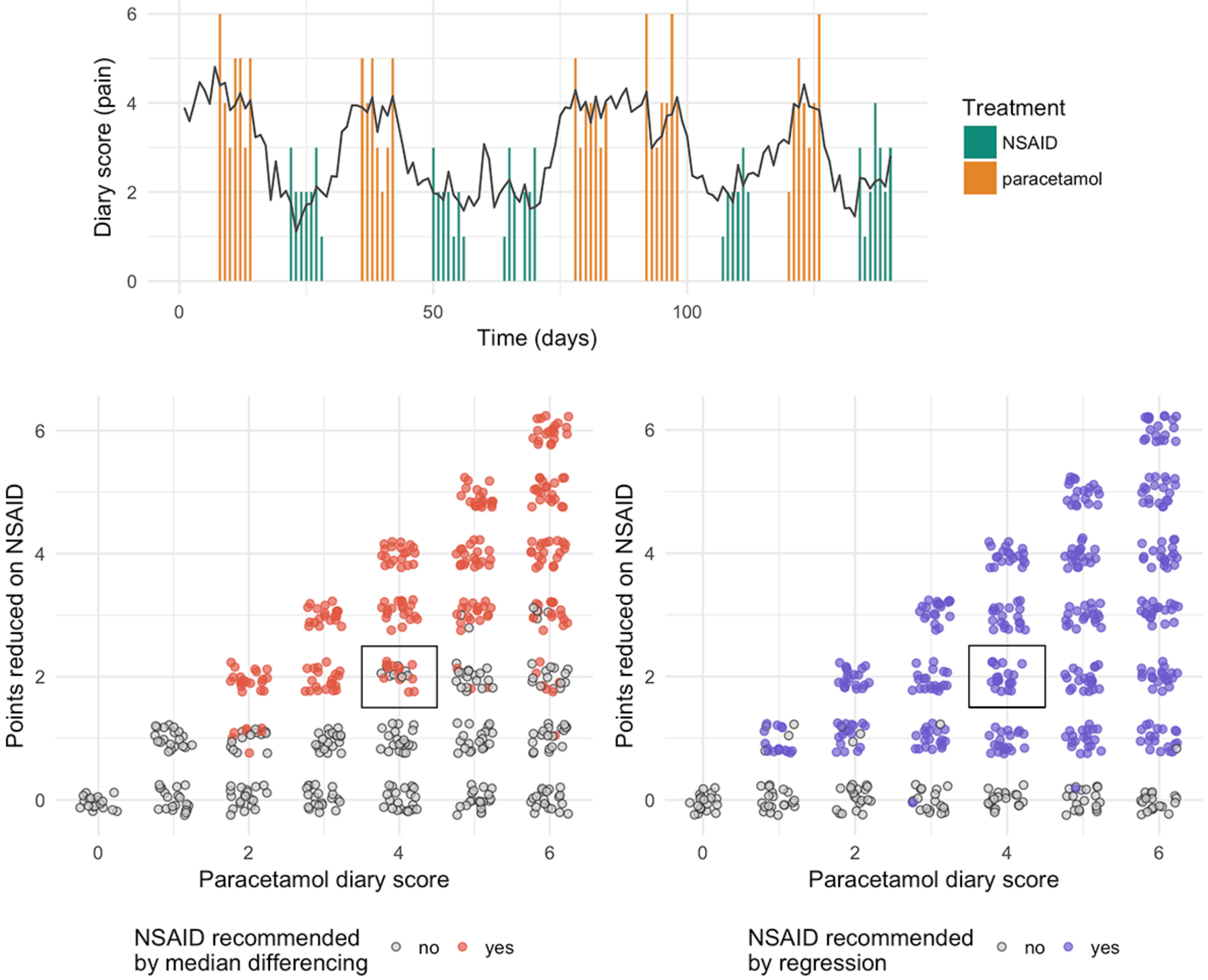


Pain Management Case Study

The trial design used in this case study emulated the design described in a study by Wegman et al [13]. Although we did not have access to the raw data for this trial and had to estimate reasonable noise parameters and wash-in/wash-out time constants, our goal was simply to compare the analysis technique from the paper with a more traditional approach involving a regression model with fixed effects for treatment and blocks [11]. The regression model we chose was the same as for the first case study.

The parameters we chose for this model can be found in Figure 3. We based our decisions about the wash-in and wash-out parameters (τ and γ) on the fact that the authors chose a wash-out period of 1 week for the different treatments and the fact that both nonsteroidal anti-inflammatory drugs (NSAIDs) and paracetamol are short-acting drugs. We converted the numeric value of the patient state to a discrete score by rounding and truncating it as shown in Table 2.

Figure 3. Analyzing a published N-of-1 study comparing NSAIDs to paracetamol. (top) An example simulation in which the true diary score on the NSAID is 2 and on paracetamol is 4. The black line shows the simulated mean outcome (unobserved) at each timepoint, and the colored bars show the observed data, which are discrete scores between 0 and 6. (bottom) A comparison of median differencing, the analysis method described in the paper, with a standard regression model. At the noise levels and effect sizes shown in (top), median differencing will recommend an NSAID only about 60% of the time (black rectangle), whereas a regression model will recommend it 100% of the time. Model parameters: tau1=tau2=1.0 day, gamma1=gamma2=3.5 days, alpha=1.0, sigma_b=0.0 (no baseline drift), sigma_p=0.5, sigma_o=1.0. NSAID: nonsteroidal antiinflammatory drug.



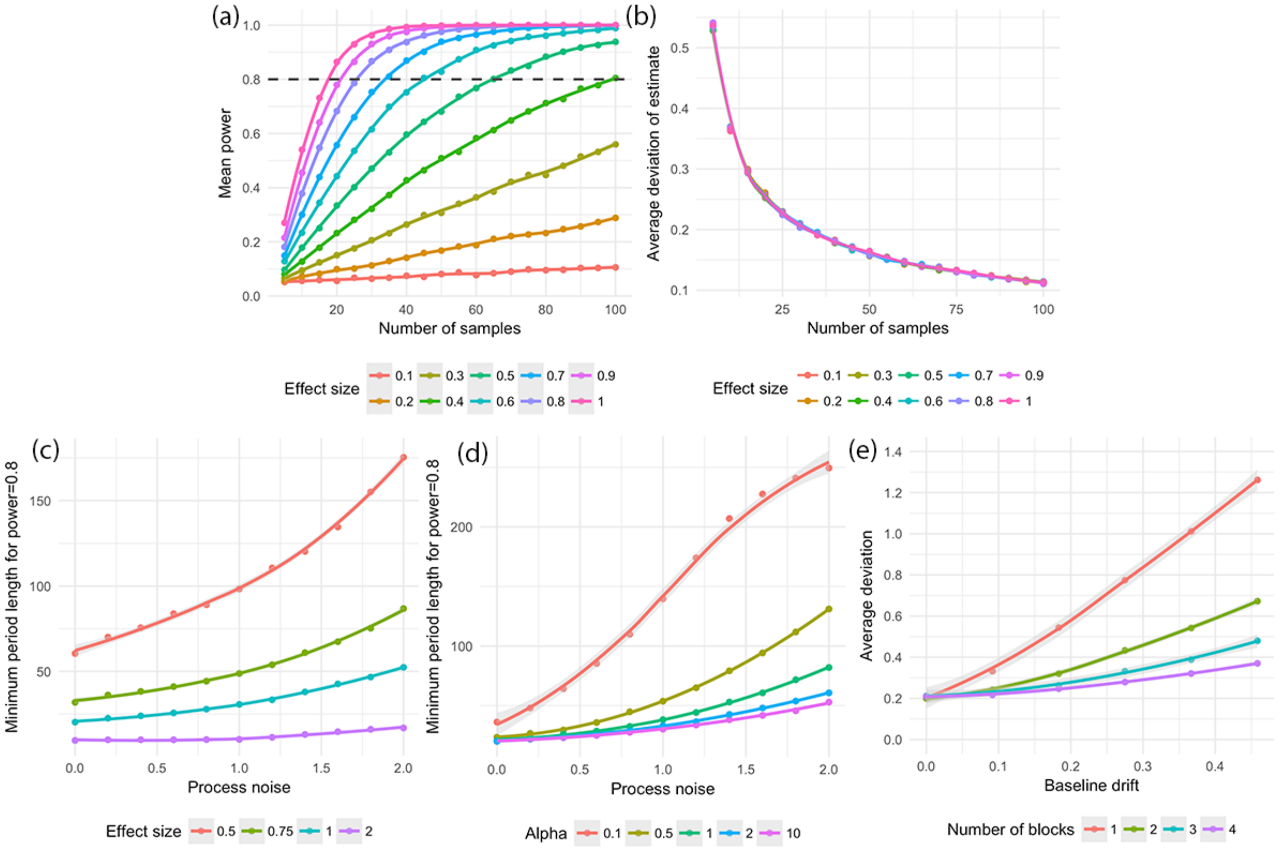
Simulations for Design Recommendations

All of the simulations in Figure 4 use a baseline of 0 and time constants (τ_1, τ_2, γ_1, and γ_2) of 0.01. Since treatment 1 is assumed to be placebo, its effect size, E_1, is 0. We used a high value for the “sensitivity to treatment effect” parameter (α=10) to produce a near-instantaneous effect. The first and second experiments in Figure 4 used only a single block, as in the absence of any sources of noise except observation noise, block design does not matter. The rest of the parameter choices are outlined in the figure. Each dot represents an average of 50 trials. The smoothed lines shown in Figure 4 are LOESS (LOcally-Estimated Scatterplot Smoothing) fits produced using geom_smooth with default parameters in ggplot, with spans of 0.4, 0.3, 1.0, 1.0, and 1.0 for subfigures a, b, c, d, and e, respectively.

Data and Code Availability

The simulation software is available in the *n1-simulator* repository under the *HD2i* organization on GitHub. Full details of the available experiments and associated plots are included with the software, along with the data sets generated in the course of making the figures.

Figure 4. Examining the effect of study design choices on power and accuracy of effect size estimates for an N-of-1 study with effectively instantaneous transitions between treatment states. (a) Effect size vs power for fixed observation noise (sigma_0=1.0) and no process noise or baseline drift. (b) Average deviation of estimate from true value vs. effect size for fixed observation noise (sigma_0=1.0) and no process noise or baseline drift. (c) Minimum treatment period length (ie. number of samples per treatment, with sampling rate fixed at 1 sample per time unit) required to attain a power of 0.8, for varying degrees of process noise and varying effect sizes. No observation noise or baseline drift is present. (d) Same as (c) except effect size is fixed at 1.0 and alpha (individual treatment response) is varied. (e) Average deviation of effect size estimate from its true value, as a function of baseline drift and number of blocks. The effect of baseline drift on the estimate is much more pronounced when fewer blocks are used. **Editorial Notice:** in (a) and (b), x-axis labels should correctly read “Number of samples per treatment.”



Results

Modeling the Key Features of an N-of-1 Study

The complete set of parameters for our model can be found in Table 1. The basic model comprises an underlying deterministic process (the growth and decay of treatment effects over time) in addition to 3 types of noise: random baseline drift (eg, long-term illness onset and recovery processes, gaining/losing weight, long-term changes in blood pressure), process noise, which manifests as short-term fluctuations (eg, heart rate and blood pressure volatility, periods of activity/inactivity, and changes in sleep and diet from day to day), and observation noise, which is a function of the instrument and is not related to any underlying biological effect (eg, the measurement noise associated with the cuff that is used to monitor blood pressure).

We divided the parameters into 4 groups: *study design parameters*, which the study designer can vary, *treatment parameters*, which are immutable features of the particular treatments under consideration, a *measurement parameter*, which is a feature of the device used to measure the outcome, and *outcome parameters*, which are features of the underlying biological process under consideration and may vary from individual to individual. A diagram of an N-of-1 block design and our model of how treatment effects vary over time is shown in Figure 1.

Case Study: Optimizing Study Design

Simulation allows us to investigate the impact of subtle design choices on the likelihood of study success. To illustrate this, we simulated a study of 2 different blood pressure medications and their impact on systolic blood pressure, similar to the data shown in [5] (see the Methods section for details). The study parameters, underlying (unobserved) data, and observed data are shown in Figure 1. The results of several hundred simulations of this study are shown in Figure 2. We used one of the standard N-of-1 regression models outlined in [6] and [11] to estimate treatment effect and obtain an associated *P* value.

In Figure 2, we see that the ordering of treatment periods has a strong effect on both statistical power and effect size estimates. On the basis of these 50 simulations, when treatments are administered in the order 1 2 1 2, power (at a standard 5% significance level) is 0.62, for 1 2 2 1 it is 0.82, for 2 1 1 2 it is 1.00, and for 2 1 2 1 it is 0.98. The median effect size estimate is also impacted by treatment ordering: for 1 2 1 2 it is 5.8, for 1 2 2 1 it is 6.6, for 2 1 1 2 it is 11.2, and for 2 1 2 1 it is 12.0. The true effect size is 10.0. We observe lower power and diminished effect size estimates for treatment orderings 1 2 1 2 and 1 2 2 1 relative to 2 1 1 2 and 2 1 2 1 as Treatment 1 takes longer to reach its full effect than Treatment 2, and the patient starts at a relatively high baseline (systolic blood pressure=160); therefore, when it is administered first, Treatment 1 never attains its full effect during the first treatment period before the transition to Treatment 2 takes place.

In Figure 2, we see the effect of sampling frequency on study power. Increasing the sampling frequency causes power to increase but only to a point. On the basis of these 50 simulations, when only 1 sample is taken at the end of each treatment period (sampling interval of 30 days), which is the most common approach to analyzing N-of-1 studies [6,11], power is only 0.14. Sampling every day (sampling interval of 1 day) yields a power of 0.84; sampling every 2 days yields a power of 0.74, every 5 days yields a power of 0.76, every 10 days yields a power of 0.56, and every 15 days yields a power of 0.50. On the basis of these results, it appears that sampling every 2 or 5 days could substantially reduce patient burden while causing only a modest reduction in power.

Figure 2 shows the effect of treatment period length, keeping the total number of blocks fixed at 2 and the sampling rate fixed at 1 sample per day. On the basis of these 50 simulations, when the treatment period length is 2 days, power is 0.18 and the mean effect size estimate is –1.5. For a period length of 5 days, power is 0.54 and the mean effect size is 3.1. For a period length of 15 days, power is 0.44 and the mean effect size is 9.7. For a period length of 30 days, power is 0.94 and the mean effect size is 10.2. For period lengths of 40, 60, and 120 days, power and mean effect sizes are 0.92 and 8.3, 0.98 and 9.7, and 0.96 and 10.6, respectively. This indicates that for a period length of 30 days, one obtains approximately as accurate an effect estimate as a period length of 60 days while shrinking the total study duration from 240 to 120 days. Period lengths that are too long run the risk of higher variance in estimates because of baseline drift, as we see with a period length of 120 days in Figure 2.

Finally, Figure 2 shows the effect of different block designs for a study of fixed length (120 days). On the basis of these 50 simulations, power for 1, 2, 3, 4, 5, and 6 blocks is 0.74, 0.86, 0.78, 0.84, 0.74, and 0.60, respectively. Mean and standard deviation of the effect size estimates are 9.7 (5.8), 9.8 (3.8), 8.7 (3.6), 8.3 (2.9), 7.0 (2.5), and 6.6 (1.8), respectively. Using 2-4 blocks appears to be the best approach, as this reduces variance in the effect size estimate relative to a single-block study. Adding more than 4 blocks increases the impact of wash-in/carryover effects on the estimate, which deviates further from its true value of 10 with each additional block.

Case Study: Evaluating Analysis Protocols

Simulation can also help us evaluate the likely success of new analysis protocols and decision criteria for N-of-1 studies. We simulated a previously published study [13] in which the outcome was a "diary score" on a scale of 0 to 6, with 0 representing "no complaints at all" and 6 representing "unbearable complaints." The study design used 5 blocks, each with 2 treatment periods; only data from the last week of each treatment period were analyzed.

In this paper, the data were analyzed as follows: the researchers took differences in median diary scores between NSAID and paracetamol treatment periods in each block and then calculated the number of treatment blocks for which the NSAID score was at least one point lower than the paracetamol score for the patient's main complaint. An NSAID was recommended if this was true in at least 4 out of 5 blocks. We refer to this method as *median differencing* from now on.

We compared median differencing to the same regression model used in the previous section [11]. Simulations show that median differencing is much more conservative in recommending an NSAID than a standard regression model trained on the same data (Figure 3). For a true effect difference of size 2 (NSAID reduces pain by 2 points relative to paracetamol), median differencing will only recommend an NSAID, on average, 61% of the time, compared with 100% of the time for the regression model. In addition, median differencing will recommend an NSAID more frequently in cases where the diary score on paracetamol is already low (the patient is not in much pain); when the score is high, it becomes harder for it to detect an effect. For a patient with a paracetamol diary score of 6 (the maximum possible pain), if the NSAID reduces the diary score to 4, median differencing will only recommend an NSAID 30% of the time, as opposed to 100% of the time for the regression model. The difference between the models is even more pronounced when the NSAID only reduces the pain score by 1; in that case, median differencing will only recommend an NSAID, on average, 7% of the time, as opposed to 92% of the time for the regression model.

Design Considerations for N-of-1 Studies

Figure 4 shows the results of a set of simulations on the basis of *best-case scenarios* — no variation in parameters other than those under investigation, as well as instantaneous treatment effects (ie, no carryover effects). The technical details of the simulations can be found in the Methods section. All of the graphs in Figure 4 relate the study design parameters to (1) statistical power—the ability to detect a treatment effect difference if it exists, and (2) the accuracy of the effect size estimates produced by the model. All compare a single treatment against placebo.

In Figures 4a and 4b, observation noise (σ_o) is fixed at 1.0, with no process noise or baseline drift. As a result, "effect size really describes a signal-to-noise ratio and is treatment and instrument agnostic." We observe that this ratio impacts power but not the accuracy of the effect estimate (Figure 4).

In Figure 4a, we see that for effect sizes of 0.1, 0.2, and 0.3, more than 100 samples per treatment are needed to obtain a power of 0.8 (at a standard 5% significance level). For an effect size of 0.4, at least 100 samples per treatment are needed. For effect sizes of 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, the numbers of samples per treatment needed to attain a power of 0.8 are approximately 65, 45, 35, 26, 21, and 18, respectively. Even more samples will be needed under real experimental conditions where process noise, baseline drift, and carryover effects all play a role. This indicates that unless the effect size is very high relative to the observation noise, N-of-1 studies using only a few blocks, with a single sample taken per block (the traditional approach to analyzing N-of-1 studies), will be vastly underpowered.

A separate consideration is the error in the effect size estimate, which declines monotonically with the number of samples. In Figure 4b, we see that to obtain an estimate within 0.2 σ_o of the true estimate, at least 30 samples per treatment are needed; to reach 0.1 σ_o, over 100 samples per treatment are needed.

Figure 4c shows the impact of process noise on the number of samples needed to attain a power of ≥0.8 at a 5% significance level in the absence of observation noise and baseline drift. In this figure, the intersample interval is fixed at 1 sample/time unit and the process noise is defined relative to that; σ_p=1.0 indicates that if no treatment effect were present, the variance of the Wiener process underlying the process noise would be 1 outcome unit/time unit. For an effect size of 0.5 and σ_p=0.0, 0.4, 0.8, 1.2, 1.6, 2.0, the numbers of samples per treatment needed to obtain a power of 0.8 are 61, 76, 89, 111, 135, and 176, respectively. For an effect size of 1.0, the numbers of samples per treatment needed are 20, 24, 28, 34, 43, and 53, respectively. Regardless of effect size, increasing the process noise from 1.0 to 2.0 roughly doubles the number of samples it takes to attain a power of 0.8. However, the effect is nonlinear; below σ_p≈1.0, the number of samples needed flattens out in the absence of other sources of noise.

In Figure 4d, we see the impact on study outcome of individual sensitivity to treatment. The lower the value of the treatment sensitivity parameter (α) is, the less effect changes in treatment have on the outcome relative to random fluctuations caused by process noise. We see this when we contrast the effect of increased process noise on the minimum samples required to attain a power of 0.8 at a significance level of 5% under conditions of low treatment sensitivity (α=0.1) and high treatment sensitivity (α=10.0). For σ_p=0.0, 0.4, 0.8, 1.2, 1.6, 2.0 and α=0.1, the numbers of samples per treatment required are 36, 64, 110, 174, 228, and 250, respectively. For α=10.0, the numbers of samples required are only 20, 23, 28, 34, 42, and 53, respectively.

Finally, Figure 4e shows us why we bother to have blocks at all: to guard against baseline drift. The figure shows what happens in a study of a total length of 240 days when block designs incorporating 1, 2, 3, or 4 blocks are used. As baseline drift increases (holding process and observation noise constant at σ_p= σ_o=0.0), the effect size estimate provided by the model increasingly deviates from its true value. This effect is most pronounced in studies with only a single block and decreases as the number of blocks increases. For example, for only 1 block, with σ_b=0.00, 0.09, 0.18, 0.27, 0.37, and 0.46, the average deviation of the effect size estimate from the true value is 0.21, 0.33, 0.54, 0.77, 1.01, and 1.26, respectively. However, with 4 blocks, with the same progression of σ_b values, the average deviation of the effect size estimate is 0.21, 0.22, 0.25, 0.28, 0.32, and 0.37, respectively.

Discussion

Summary of the Paper

We have developed a stochastic time-series model that simulates an N-of-1 study, facilitating rapid optimization of N-of-1 study designs and increasing the likelihood of study success while minimizing participant burden. We have used this model to evaluate 2 case studies, showing how the number of treatment blocks, ordering of treatments within blocks, duration of each treatment, sampling frequency, and study analysis protocol affect our ability to detect true differences in treatment efficacy. Our simulation software is available on GitHub as described in the Methods section.

Recommendations for the Design of N-of-1 Studies

An N-of-1 study should have as many blocks as possible to avoid baseline drift (Figure 4). If no wash-in or carryover effects are present, a single sample should be taken at the end of each of *JN* different treatment periods, where *N* is the number of blocks and *J* the number of treatments; *N* should be made as high as possible; each block should be made as short as possible. However, in practice, the number of blocks we can use in a study is bounded by the dangers of administering different treatments in rapid succession, the time it takes treatments to ramp up to their full effects (“run-in”: Table 1), the time it takes them to stop working when they are discontinued (“wash-out”: Table 1), and participant patience.

It is important to consider the fact that most N-of-1 studies of reasonable length and reasonable sampling frequency will be underpowered unless the difference in treatment effects is at least on the order of the standard deviation of the observation noise (Figure 4). The goal, perhaps obvious, should be to measure the outcome with as little noise as possible and at as high a frequency as possible, and/or to continue the study until enough samples are obtained to ensure that the effect will be detected if it is there.

Finally, it is important to remember the difference between power and accuracy. Just because a statistically significant difference in treatment effects is detected, it does not mean that the quantitative estimate of E_2-E_1 reported by the model is accurate. Even when a study is sufficiently powered, the effect size estimate will almost always improve with the addition of more samples.

Beyond these general statements, our main recommendation for N-of-1 study designers is to simulate the study. We can see from Figures 4c and d that process noise and individual sensitivity to treatment can have a dramatic impact on the number of samples needed to adequately power a study, especially if the effect size is small. The choice of analysis method can also have a substantial impact on study outcome and treatment recommendations (Figure 3); therefore, it is important to compare novel analysis methods to the standard models provided by the AHRQ and others [6,11]. Simulations can help in both cases.

Modeling Different Outcome Types

Most of our analyses in this paper concerned a continuous (or near-continuous) random variable, such as blood pressure or heart rate. However, many N-of-1 trials examine outcomes that are better modeled as counts, proportions, binary random variables (yes/no), or discrete bounded scores (such as surveys). Studies with these outcome types can be simulated by transforming the output of the stochastic differential equation model using a set of transformations similar to those for generalized linear models (see Table 2). We used one such transformation to discretize the scores for the pain management case study.

Sources of Treatment and Instrument Parameters

By far, the strongest drawback to the simulation approach is the difficulty associated with identifying reasonable simulation parameters, especially in cases where the outcome is not a continuous value (see Table 2).

Some parameters have relatively clear interpretations and can be found by looking at the known characteristics of treatments and instruments. For example, in the case of a continuous-valued outcome, we can think of the treatment effect, $X_j(t)$, as the treatment's maximum impact—at each point in time—on the outcome in the absence of any noise, in a population of people exactly like the one who is undergoing the study. The treatment effect is governed by 3 parameters: τ_j, the time constant of "wash-in" for that treatment, γ_j, the time constant of "wash-out", and E_j, the asymptotic effect size (the change from baseline that the person would experience in the long run was he/she to continue on this treatment). In the case of a pharmaceutical intervention, these are important parameters that have probably been estimated in earlier clinical trials and used to guide dosages, dosing frequencies, etc. Similarly, reasonable values for σ_o can often be obtained from technical specifications of whatever instrument is used to monitor the outcome.

The emerging field of mobile health may provide some help in estimating parameters like σ_p and σ_b, which are properties of an outcome and its natural variation over time [14]. As we begin to monitor patients longitudinally with increasingly higher resolution, our quantitative understanding of long- and short-term variation in biological processes will naturally increase. However, in simulations at present, we recommend experimenting with varying parameter scales and examining raw plots of the data to see if the level of noise produced by the model is reasonable. It may also make sense to test ranges of α, σ_b, and σ_p and examine plots like those shown in Figure 4 to assess the effect of these parameter choices on statistical models.

Study Limitations and Future Work

This study fits simulated data with a simple regression model recommended by the AHRQ, but the data themselves are simulated using a more realistic model. A natural next step would be to use the full simulation model as the basis for fitting data. Future versions of our software will allow users to fit data using the AHRQ model and the full time-series model in a Bayesian framework, which infers the model parameters using posterior probability distributions given the data rather than point estimates [15,16]. Thus, uncertainty is an inherent part of the model. This will provide a basis for directly comparing the performance of the full time-series model against the simple AHRQ model for making treatment recommendations. In addition, posterior parameter distributions inferred from real data can be used to generate more realistic simulated data. This will be especially useful for studies with discrete outcomes, where the linkage between model parameters and outcome data is more difficult to interpret. Another advantage of a Bayesian parameter estimation approach is that it allows parameter estimates for N-of-1 studies to be continually updated as more individuals undergo the same study, creating a system that learns from past data to adapt the design of future studies.

One important limitation of our model is that although it incorporates multiple sources of noise, it ignores more structured sources of outcome variation (eg, variation in heart rate does not principally happen stochastically with time, but the heart rate does show structured change across hours, days, and ovulatory cycles). It is also possible that long-term seasonal, day of week, and time of day effects can influence the outcome of N-of-1 studies. Future versions of our model may incorporate parameters for these effects and fit them using methods akin to those of Prophet [17] or other Bayesian time-series models. In the meantime, users can address these issues by manually adding known sources of variation to the baseline drift term or by choosing outcome parameters that "average out" known sources of variation (eg "heart rate daily mean").

In general, the development of realistic simulations of N-of-1 studies is an ongoing process. We believe that simulation will prove crucial as N-of-1 studies enter mainstream clinical practice, especially in the realm of precision medicine, and we hope that our model will inspire others to adopt N-of-1 studies as a tool in their own research.

Authors' Contributions

BP and NZ jointly conceived of the idea for an N-of-1 simulation model. BP and EBB designed the model, wrote the model code, and conducted the experiments for the paper. EBB translated the code into R and created the documentation and user-friendly interface. BP drafted the manuscript. MJ, JTD, and NZ provided extensive feedback on the manuscript and model design.

Conflicts of Interest

None declared.

References

1. Duan N, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. J Clin Epidemiol 2013 Aug;66(8 Suppl):S21-S28 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.04.006] [Medline: 23849149]
2. Gabler N, Duan N, Vohra S, Kravitz R. N-of-1 trials in the medical literature: a systematic review. Med Care 2011 Aug;49(8):761-768. [doi: 10.1097/MLR.0b013e318215d90d] [Medline: 21478771]
3. Kravitz RL, Duan N, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. Milbank Q 2008 Dec;86(4):533-555 [FREE Full text] [doi: 10.1111/j.1468-0009.2008.00533.x] [Medline: 19120979]
4. Kravitz R, Paterniti D, Hay M, Subramanian S, Dean D, Weisner T, et al. Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. Contemp Clin Trials 2009 Sep;30(5):436-445. [doi: 10.1016/j.cct.2009.04.001] [Medline: 19375521]
5. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Per Med 2011 Mar;8(2):161-173 [FREE Full text] [doi: 10.2217/pme.11.7] [Medline: 21695041]
6. Kravitz R, Duan N, Eslick I, Gabler N, Kaplan H, Kravitz R, et al. Agency for Healthcare Research and Quality. Rockville, MD; 2014. Design and Implementation of N-of-1 Trials: A User's Guide URL: https://effectivehealthcare.ahrq.gov/topics/n-1-trials/research-2014-5 [accessed 2019-02-04] [WebCite Cache ID 75wTPSSaw]
7. Guyatt G, Keller J, Jaeschke R, Rosenbloom D, Adachi J, Newhouse M. The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. Ann Intern Med 1990 Feb 15;112(4):293-299. [doi: 10.7326/0003-4819-112-4-293] [Medline: 2297206]
8. Zucker D, Schmid C, McIntosh M, D'Agostino RB, Selker H, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. J Clin Epidemiol 1997 Apr;50(4):401-410. [doi: 10.1016/S0895-4356(96)00429-5] [Medline: 9179098]
9. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. J Clin Epidemiol 2010 Dec;63(12):1312-1323 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.020] [Medline: 20863658]
10. Nikles J, Mitchell GK, Schluter P, Good P, Hardy J, Rowett D, et al. Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. J Clin Epidemiol 2011 May;64(5):471-480. [doi: 10.1016/j.jclinepi.2010.05.009] [Medline: 20933365]
11. Mengersen K, McGree J. Statistical analysis of N-of-1 trials. In: The Essential Guide to N-of-1 Trials in Health. Dordrecht: Springer; 2015:135-153.
12. Holford N, Kimko HC, Monteleone JP, Peck CC. Simulation of clinical trials. Annu Rev Pharmacol Toxicol 2000;40:209-234. [doi: 10.1146/annurev.pharmtox.40.1.209] [Medline: 10836134]
13. Wegman A, van der Windt DA, de Haan M, Devillé WL, Fo CT, de Vries TP. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Ann Rheum Dis 2003 Dec;62(12):1156-1161 [FREE Full text] [doi: 10.1136/ard.2002.002865] [Medline: 14644852]
14. Steinhubl S, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
15. Hoffman MD, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014;15(1):1593-1623 [FREE Full text]
16. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Softw 2017;76(1):1-32. [doi: 10.18637/jss.v076.i01]
17. Taylor SJ, Letham B. Forecasting at Scale. The American Statistician 2018;72(1):37-45 [FREE Full text] [doi: 10.1080/00031305.2017.1380080]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
NSAID: nonsteroidal anti-inflammatory drug
LOESS: locally-estimated scatterplot smoothing

Edited by G Eysenbach; submitted 29.10.18; peer-reviewed by B Smarr, L Zhang; comments to author 28.11.18; revised version received 28.12.18; accepted 29.12.18; published 01.04.19.

Please cite as:
Percha B, Baskerville EB, Johnson M, Dudley JT, Zimmerman N
Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations
J Med Internet Res 2019;21(4):e12641
URL: https://www.jmir.org/2019/4/e12641/
doi:10.2196/12641
PMID:30932871

Original Paper

Wearable Proximity Sensors for Monitoring a Mass Casualty Incident Exercise: Feasibility Study

Laura Ozella[1], PhD; Laetitia Gauvin[1], PhD; Luca Carenzo[2,3], MD; Marco Quaggiotto[1,4], PhD; Pier Luigi Ingrassia[3], PhD; Michele Tizzoni[1], PhD; André Panisson[1], PhD; Davide Colombo[2], PhD; Anna Sapienza[1,5], PhD; Kyriaki Kalimeri[1], PhD; Francesco Della Corte[2], MD; Ciro Cattuto[1], PhD

[1]Data Science Laboratory, Institute for Scientific Interchange Foundation, Torino, Italy

[2]Department of Translational Medicine, Eastern Piedmont University, Novara, Italy

[3]Centro Interdipartimentale di Didattica Innovativa e di Simulazione in Medicina e Professioni Sanitarie SIMNOVA, Universita` del Piemonte Orientale, Novara, Italy

[4]Department of Design, Politecnico di Milano, Milano, Italy

[5]University of Southern California Information Sciences Institute, Marina del Rey, CA, United States

Corresponding Author:
Laura Ozella, PhD
Data Science Laboratory
Institute for Scientific Interchange Foundation
Via Chisola 5
Torino, 10131
Italy
Phone: 39 3491973277
Email: laura.ozella@gmail.com

Abstract

Background: Over the past several decades, naturally occurring and man-made mass casualty incidents (MCIs) have increased in frequency and number worldwide. To test the impact of such events on medical resources, simulations can provide a safe, controlled setting while replicating the chaotic environment typical of an actual disaster. A standardized method to collect and analyze data from mass casualty exercises is needed to assess preparedness and performance of the health care staff involved.

Objective: In this study, we aimed to assess the feasibility of using wearable proximity sensors to measure proximity events during an MCI simulation. In the first instance, our objective was to demonstrate how proximity sensors can collect spatial and temporal information about the interactions between medical staff and patients during an MCI exercise in a quasi-autonomous way. In addition, we assessed how the deployment of this technology could help improve future simulations by analyzing the flow of patients in the hospital.

Methods: Data were obtained and collected through the deployment of wearable proximity sensors during an MCI functional exercise. The scenario included 2 areas: the accident site and the Advanced Medical Post, and the exercise lasted 3 hours. A total of 238 participants were involved in the exercise and classified in categories according to their role: 14 medical doctors, 16 nurses, 134 victims, 47 Emergency Medical Services staff members, and 27 health care assistants and other hospital support staff. Each victim was assigned a score related to the severity of his/her injury. Each participant wore a proximity sensor, and in addition, 30 fixed devices were placed in the field hospital.

Results: The contact networks show a heterogeneous distribution of the cumulative time spent in proximity by the participants. We obtained contact matrices based on the cumulative time spent in proximity between the victims and rescuers. Our results showed that the time spent in proximity by the health care teams with the victims is related to the severity of the patient's injury. The analysis of patients' flow showed that the presence of patients in the rooms of the hospital is consistent with the triage code and diagnosis, and no obvious bottlenecks were found.

Conclusions: Our study shows the feasibility of the use of wearable sensors for tracking close contacts among individuals during an MCI simulation. It represents, to our knowledge, the first example of unsupervised data collection—ie, without the need for the involvement of observers, which could compromise the realism of the exercise—of face-to-face contacts during an MCI exercise. Moreover, by permitting detailed data collection about the simulation, such as data related to the flow of patients in the hospital, such deployment provides highly relevant input for the improvement of MCI resource allocation and management.

(J Med Internet Res 2019;21(4):e12251) doi:10.2196/12251

KEYWORDS

contact patterns; contact networks; wearable proximity sensors; mass casualty incident; simulation; medical staff – patient interaction; patients' flow

Introduction

Background

A mass casualty incident (MCI) is defined as a situation in which, at a certain time, the available care resources are unable to meet the demand for medical care of the incident [1]. Each year, MCIs occur worldwide and are caused by conventional causes such as weapons, explosions, vehicular and airplane accidents, and deliberate or spontaneous chemical mass intoxications. These incidents require emergency health care teams to treat large numbers of injured victims [2], and this might compromise the normal functioning of hospitals. Simulation applied to health care is rapidly gaining acceptance in medical and academic communities, and it can be a valuable tool for better training for management of MCIs. It provides a safe, controlled environment in which it is possible to test plans and procedures and improve them, as well as to evaluate policies and guidelines [3]. Traditionally, actors are used in disaster exercises, and they are coached to mimic and exhibit realistic manifestations of several medical and traumatic pathologic states that may be present in a real MCI [2-4]. Although the use of simulation in medical education has increased over the last 2 decades, collecting and analyzing data of a mass casualty functional exercise still occurs in an unsystematic manner, without a standardized method. Common methods to assess performances during MCI simulations are direct observations of functional exercise performance and video analysis of participants' behaviors [5]. However, these methods present some limitations: the focus of the observers' attention can subjectively vary, and the need of several observers, both for direct observations and for videos, could affect the realism of the event and decrease the level of emotional engagement of the participants [3]. An objective and reproducible method that identifies the strengths and weaknesses of simulations is required to lead the improvement in the response system [6]. In mass casualty simulations, wireless medical sensor networks and Radio-Frequency IDentification (RFID) technology have been used to track information about the status of the casualties, thus providing timely situational awareness during exercises [3,7,8], and the use of RFID was compared with manual data collection, demonstrating the reliability and applicability of the system [3]. Wearable proximity devices could provide not only patient information and tracking capability to locate people and equipment but also information on interactions among individuals. Wearable sensors have been successfully used to measure face-to-face proximity relations in various hospital settings that include the pediatric ward [9] and acute care geriatric unit [10,11].

Objectives

The use of proximity sensors in the field of MCI simulation could provide a continuous and fully distributed collection system of high-resolution data on the interactions among patients and medical staff to investigate the dynamics of interactions with regard to the different roles and severity of the patient's injuries. In this study, we have illustrated the feasibility of contact measures through wearable proximity sensors in a live MCI simulation aimed at providing data-driven knowledge to perform debriefing and identify room for improvement. The main objectives of our study were (1) to investigate the interactions between medical staff and patients with regard to their roles and severity of the victim's condition and (2) to estimate the presence of victims in different spaces of the field hospital to study the patients' flow.

Methods

Study Setting

A building collapse following a flood was simulated during an MCI functional exercise organized in Novara, Italy, on May 19, 2016, from 7 pm to 10:30 pm. The MCI exercise was organized in the framework of the residential course of the European Master in Disaster Medicine (EMDM). The EMDM is an international 12-month–long blended learning master's degree program for health care providers involved in medical preparedness and response to disasters [12]. The exercise included both a prehospital and an in-hospital disaster response phase. The scenario comprised 2 locations: the building collapse site (prehospital response) and the field hospital (in-hospital response). The hospital was located approximately 2 kilometers from the incident site.

Overall, exercise participants were distinguished into the following classes for the purpose of this study and based on their role in the simulation: Medical Doctors (*MD*), Nurses (*Nurse*), Emergency Medical Service (*EMS*) personnel, and Health Care Assistants (*HCA*) and simulated victims (*Victim*). Victims were portrayed by medical students. They attended an introductory course on disaster medicine (8-hour live lectures) and specific training on how to simulate clinical conditions provided in an individual victim storyboard, when to change the dynamic casualty cards (DCCs) reporting their vitals according to the treatment applied, and how to properly collect data (2-hour live lectures). Details about the casualty evolution method, general structure of the simulation and DCCs were described in a series of previous papers [3,6,13,14]. The EMDM students acted as doctors and nurses and were distributed as follows: 8 physicians and 5 nurses staffed the ambulances provided by local EMS agencies as prehospital response, whereas 6 physicians and 11 nurses were in the field hospital that had been previously deployed by the Italian Army. EMS personnel and HCAs were played by local ambulance volunteers (basic emergency medical technician level) and first aid–trained soldiers, respectively. None of the participants had been previously informed about the scenario.

Expected Triage and Injury Severity Score

According to their predetermined storyboard, each victim had an expected initial triage category according to the Simple Triage and Rapid Treatment protocol [15]: 6 victims were *Black*, 15 were *Red*, 27 were *Yellow*, and 86 were *Green*. Responders had to assign a triage score to each victim during the exercise, both at the accident site (*on-scene triage*) and at the hospital (*hospital triage*). In Section 2 of the Multimedia Appendix 1, we report the final triage accuracy of the exercise.

Victims (*Black* group excluded) were also classified based on their injuries using the Injury Severity Score (ISS) [16]. The ISS is an established medical score to assess trauma severity with a range from 1 to 75, grouped by 5 categories: *Minor* (1-3), *Moderate* (4-8), *Serious* (9-15), *Severe* (16-24), and *Critical* (25-75). In addition, the category *Nontraumatic* that indicates victims without physical trauma (such as anxiety crises) was added. In total, 21 victims were assigned to the *Nontraumatic* group, 36 to *Minor* group, 47 to *Moderate* group, 9 to *Serious* group, 4 to *Severe* group, and 11 to *Critical* group.

Data Collection

Data collection was performed as described below. Each participant wore a wearable proximity sensor: the sensor was inserted into a transparent envelope and fixed with adhesive tape at the center of the chest (on the sternum area) to detect person-to-person interactions. At the beginning of the simulation, victims were both at the accident site and at the hospital as regular in-hospital patients. During the exercise, victims could take 1 of 3 possible pathways: (1) transferred from the accident sites to the hospital by ambulances and minibuses; (2) transferred to another virtual hospital; and (3) discharged from the simulation. The exact times of the transfers as well as the ending time of the simulation for each victim were marked by external observers. In addition, proximity sensors were placed on the ceiling of the rooms (tents) of the hospital area (Figure 1; category *Location*) as fixed tags. Table 1 reports a summary of the total number of sensors for each category in the Prehospital and Hospital area.

The sensor setup was designed by the SocioPatterns collaboration consortium [17]. This system is based on wearable proximity sensors (*tags*) that exchange ultra–low power radio packets in a peer-to-peer fashion [9,18-20]. Sensors in close proximity exchange with one another a maximum of about 1 power packet per second, and the exchange of low-power radio packets is used as a proxy for the spatial proximity between tags [9,18]. In particular, close proximity is measured by the attenuation, defined as the difference between the received and transmitted power. Each device has a unique identification number that was used to link the information on the individual carrying the device with his/her profile or, in the case of fixed tags, with the location where the sensors are placed. More details on the functioning of the tags and on the data collection pipeline are given in Section 1 of the Multimedia Appendix 1.

Figure 1. A map of the field hospital.

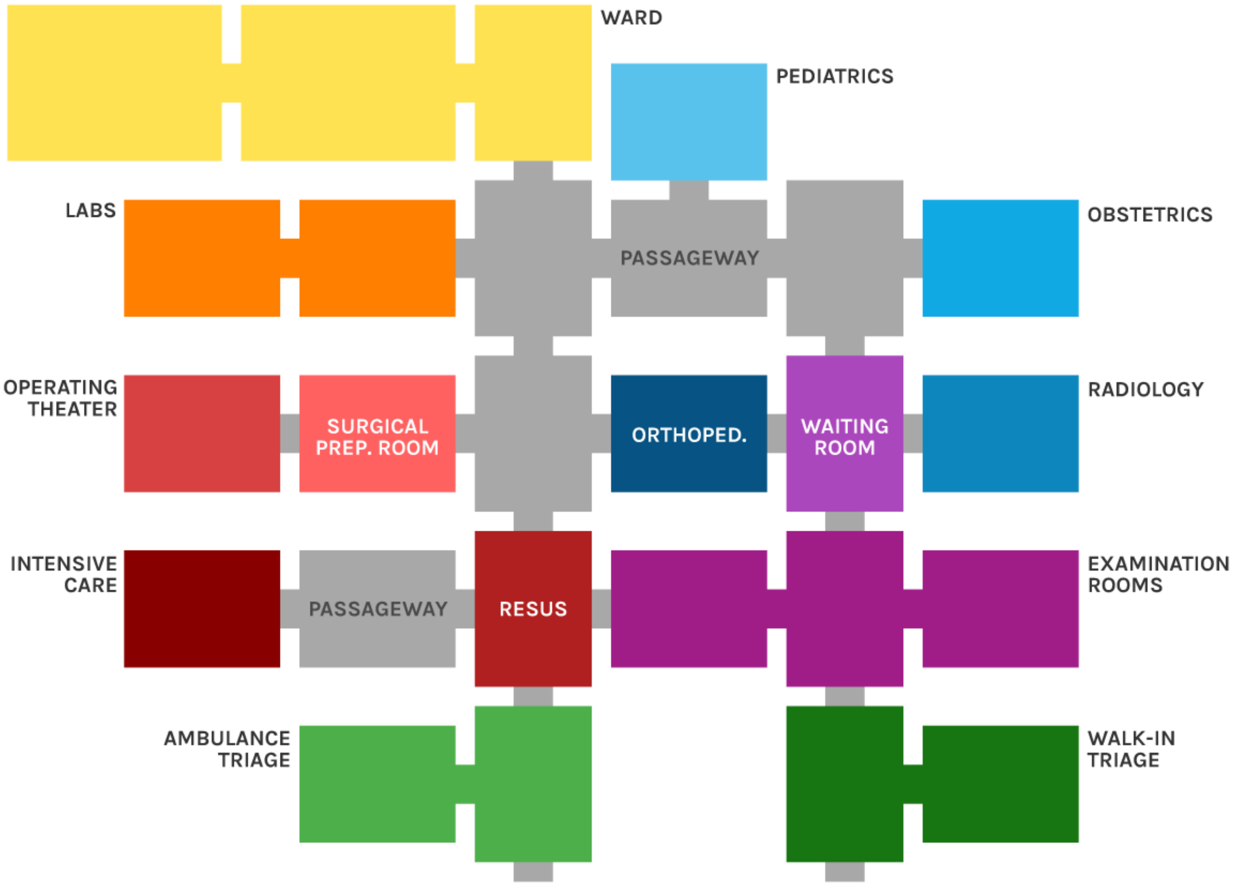


Table 1. Number of proximity sensors in the Prehospital and Hospital area by category.

Category	Prehospital	Hospital	Total
Medical Doctor	8	6	14
Nurse	5	11	16
Victim	112	22	134
Emergency Medical Service	47	—[a]	47
Health Care Assistant	—	27	27
Location	—	30	30
Total	172	96	268

[a] Not applicable.

Contacts Among Participants

We analyzed the contacts among participants belonging to categories *Victim*, *MD*, *Nurse*, and *EMS* for the Prehospital area and the contacts between participants belonging to categories *Victim*, *MD*, *Nurse*, and *HCA* for the Hospital area. We considered the contacts between individuals across categories, both in the Prehospital and Hospital area. We defined that a *contact* occurs between 2 individuals during a time slice duration of 20 seconds if the proximity devices worn by the participants exchanged at least 1 radio packet during that interval and the median attenuation of received packets exceeds an attenuation threshold of 70 dBm. After a contact is established, it is considered ongoing as long as the devices continue to exchange at least 1 such packet for every subsequent 20-second interval [18]. The system was set to detect proximity events between devices situated within 1 to 1.5 m of one another. This setting ensures that when individuals wear the devices on their chest, exchange of radio packets between devices is only possible when they are facing each other, as the human body acts as a radio-frequency shield at the carrier frequency used for communication. This system allows us to monitor the number of contacts and their duration. Data were extracted and cleaned separately for each sensor, and those collected before 7 pm and after 10:30 pm were discarded to keep track of only meaningful proximity events. Moreover, for each victim, we discarded the data collected after the exact time of simulation end (death, discharge, or end-of-simulation time). We analyzed the data separately for the Prehospital and Hospital area. With regard to the victims transferred from the accident site to Hospital during the simulation, we considered the data collected before the exact time of transfer belonging to Prehospital data and data collected after this time belonging to Hospital data.

We generated aggregated networks of contacts between participants on the full exercise duration, both in the Prehospital and Hospital area, to study the close range interactions during the exercise as well as to confront the results with those obtained in different real-world settings. We call k_i the degree of a node i, ie, the number of distinct individuals with whom individual i has been in contact, and w_{ij}, the weight of an edge between nodes i and j, ie, the cumulative duration of the contact events recorded between 2 individuals, i and j.

Then, we generated contact matrices based on the median cumulative time spent in proximity between victims with different triage and ISSs and the caregivers (medical doctors, nurses, emergency medical services, and health care assistants). Time spent in proximity with victims for each caregivers' category was compared using the Kruskal-Wallis test. We respectively considered the *on-scene triage* scores and the *hospital triage* scores to build the Prehospital and Hospital matrix (ie, the triage scores assigned by the medical doctors).

Presence of Victims in the Field Hospital

We estimated the presence of the patients in different rooms of the field hospital by analyzing the power packets exchanged between sensors worn by individuals belonging to the category *Victim* and the fixed sensors belonging to category *Location* in the Hospital area. To assess the location of a patient in a given room at a given time, we set up 2 thresholds on the count of power packets exchanged between the devices respectively to evaluate the presence of the participant in the exercise and the presence of the participant in a given room. For each time slot of 5 min, we assumed that a participant is still present in the exercise site if the total count of the power packets exchanged between all the fixed tags and the participant's tag is greater than 15. We assumed that a participant is present in a given room if the total number of power packets exchanged between his/her tag and the fixed tag of the room considered is higher than 5 for each time slot.

This allows describing the patients' flow through the rooms of the field hospital. The field hospital consisted of 27 rooms organized as follows: 3 general wards, 2 laboratories, 5 passageways (hallways), 3 examination rooms, 1 pediatric ward, 1 waiting room, 1 obstetric ward, 1 operating theatre with 2 beds, 1 surgical preparation room, 1 orthopedic ward, 1 radiology waiting room, 1 radiology, 1 intensive care with 3 beds, 1 emergency department resuscitation (resus) area, 2 ambulance triage rooms (patients brought in by ambulance), and 2 walk-in triage rooms. In this analysis, we grouped the patients' wards, the laboratories, the passageways, the examination rooms, the ambulance triage rooms, and walk-in triage rooms in the same space.

Presence Patterns of Victims in the Hospital

To study the link between the presence patterns of victims and the conditions of the victims, we used the t-Stochastic Neighbor Embedding (t-SNE) technique that converts a high-dimensional data set into a matrix of pair-wise similarities and allows to visualize the resulting similarity data [21]. In this study, we used, as an input dataset, a set of vectors where each vector described the spatial features of each victim. More exactly, each victim is initially represented as a vector where elements are time spent by that patient in a given room of the field hospital (normalized on total presence duration).

Results

Network Analysis and Contact Among Victims and Rescuers

A total of 238 individuals participated in the exercise. They were categorized as follows: 14 MD, 16 Nurses, 134 Victims, 47 EMS, and 27 HCA. The contacts within the same category were not included in this analysis. Figure 2 shows the degree and weight distribution in the Prehospital and Hospital area. The aggregated contact network in the Prehospital area is formed by 172 nodes and 2035 edges, and the average degree is k=23.66 (range 1-94), and in the Hospital area, the network is formed by 124 nodes and 1335 edges, and the average degree is k=21.53 (range 1-58). The weight distribution is heterogeneous in both areas, with heavy-tailed distributions: most contacts are short, and there are few long-lasting contacts.

Contact matrices reveal different amounts of time spent in proximity depending on the severity of the patient and the role of the caregiver (Figures 3 and 4). At the scene of the accident, there was a significant difference between the time spent in proximity between EMS and victims both with regard to the triage (X^2_3=19.479; *P*<.001) and to the ISS (X^2_5=36.106; *P*<.001). The higher time in contact was with Green victims and with victims classified as Moderate. Regarding the triage, there were no significant differences between the time spent in proximity with the victims for both MD and nurses. On the contrary, regarding the ISS, there were significant differences for MD (X^2_5=13.576; *P*=.02) and nurses (X^2_5=12.798; *P*=.02). Both categories spent higher time in contact with victims classified as Moderate and Critical. At the field hospital, there was a significant difference between the time spent in proximity between HCA and victims both with regard to the triage (X^2_3=31.271; *P*<.001) and to the ISS (X^2_5=46.989; *P*<.001). The higher time in contact was with Green victims and with victims classified as Minor and Moderate. There were no significant differences between the time spent in proximity between MD and victims (with regard to the triage and ISS) and nurses and victims (with regard to the triage). However, there was a significant difference between the time spent in proximity between nurses and victims with regard to the ISS (X^2_5=14.965; *P*=.01), nurses spent the higher time in contact with victims classified as Moderate and Serious.

Figure 5 shows the cumulative time in contact (normalized on total number of participants belonging to each caregiver category) between caregivers and victims with different triage at the Prehospital area (panel A) and Hospital area (panel B). The rescuers who spent more cumulative time in contact with victims were nurses at the scene of the accident and HCA at the field hospital.

Figure 2. Degree and weight distributions. Degree distribution P(k) of the aggregated contact networks, in the Prehospital area (panel A) and in the Hospital area (panel C). Distribution of the weights of the aggregated contact networks in the Prehospital area (panel B) and in the Hospital area (panel D).

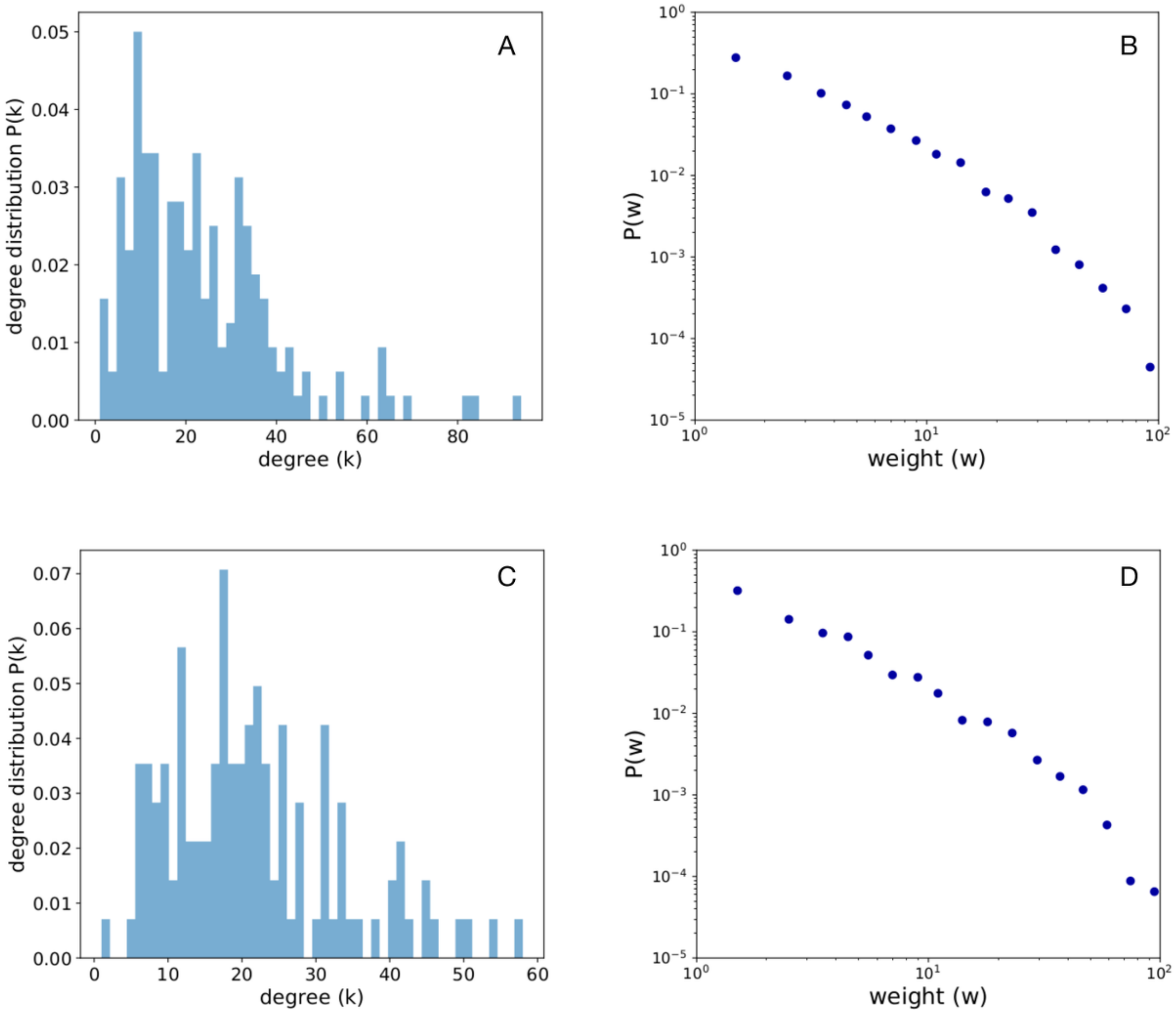
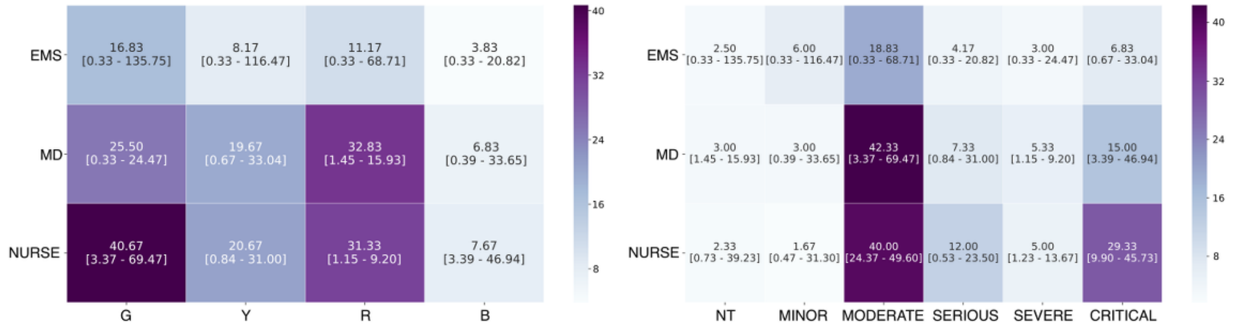


Figure 3. Prehospital contact matrices. Median of cumulative time spent (in minutes) between patients with different triage score and rescuers (left panel); Median of cumulative time spent (in minutes) between patients with different Injury Severity Scores (Nontraumatic, Minor, Moderate, Serious, Severe, and Critical) and rescuers (right panel). 95% CIs are indicated in brackets. EMS: Emergency Medical Services; MD: Medical Doctors; NT: nontraumatic; G: Green; Y: Yellow; R: Red; B: Black.



	G	Y	R	B
EMS	16.83 [0.33 - 135.75]	8.17 [0.33 - 116.47]	11.17 [0.33 - 68.71]	3.83 [0.33 - 20.82]
MD	25.50 [0.33 - 24.47]	19.67 [0.67 - 33.04]	32.83 [1.45 - 15.93]	6.83 [0.39 - 33.65]
NURSE	40.67 [3.37 - 69.47]	20.67 [0.84 - 31.00]	31.33 [1.15 - 9.20]	7.67 [3.39 - 46.94]

	NT	MINOR	MODERATE	SERIOUS	SEVERE	CRITICAL
EMS	2.50 [0.33 - 135.75]	6.00 [0.33 - 116.47]	18.83 [0.33 - 68.71]	4.17 [0.33 - 20.82]	3.00 [0.33 - 24.47]	6.83 [0.67 - 33.04]
MD	3.00 [1.45 - 15.93]	3.00 [0.39 - 33.65]	42.33 [3.37 - 69.47]	7.33 [0.84 - 31.00]	5.33 [1.15 - 9.20]	15.00 [3.39 - 46.94]
NURSE	2.33 [0.73 - 39.23]	1.67 [0.47 - 31.30]	40.00 [24.37 - 49.60]	12.00 [0.53 - 23.50]	5.00 [1.23 - 13.67]	29.33 [9.90 - 45.73]

Figure 4. Hospital contact matrices. Median of cumulative time spent (in minutes) between patients with different triage score and rescuers (left panel); Median of cumulative time spent (in minutes) between patients with different Injury Severity Scores (Nontraumatic, Minor, Moderate, Serious, Severe, and Critical) and rescuers (right panel). 95% CIs are indicated in brackets. HCA: Health Care Assistants; MD: Medical Doctors; NT: nontraumatic; G: Green; Y: Yellow; R: Red.

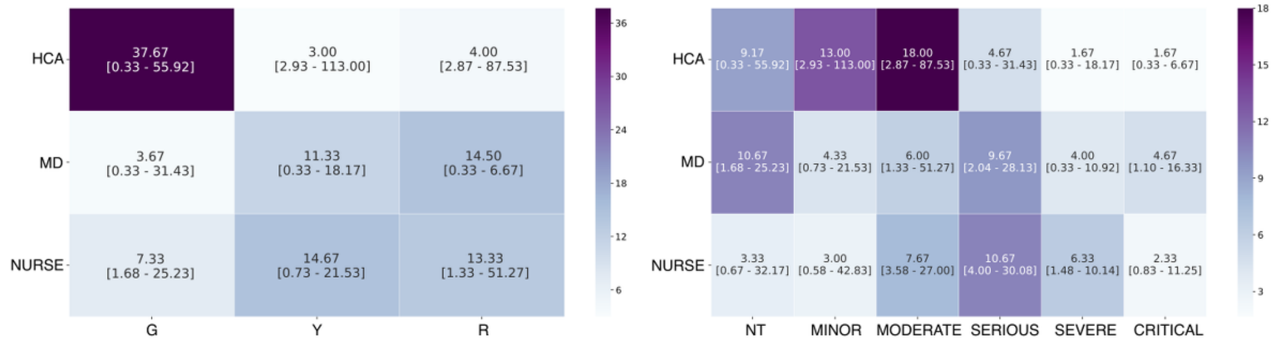
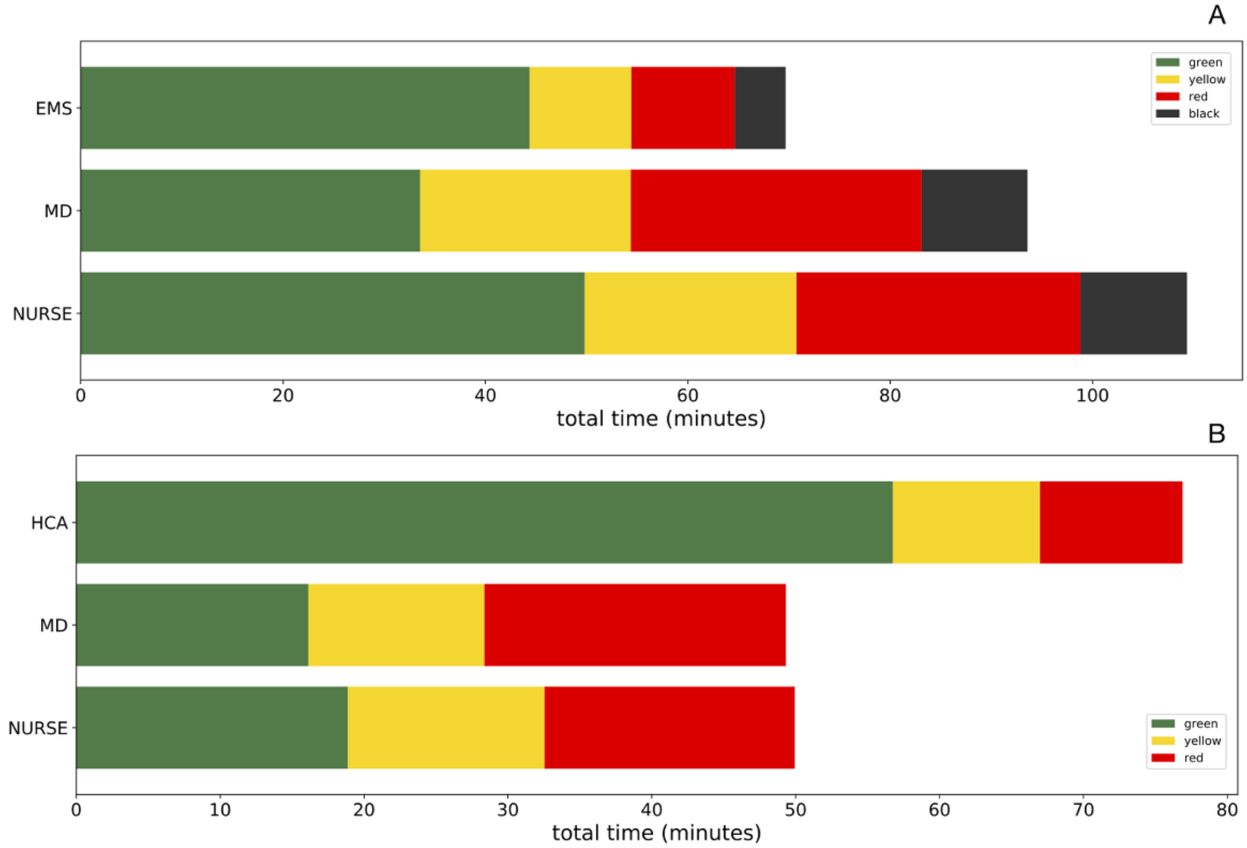


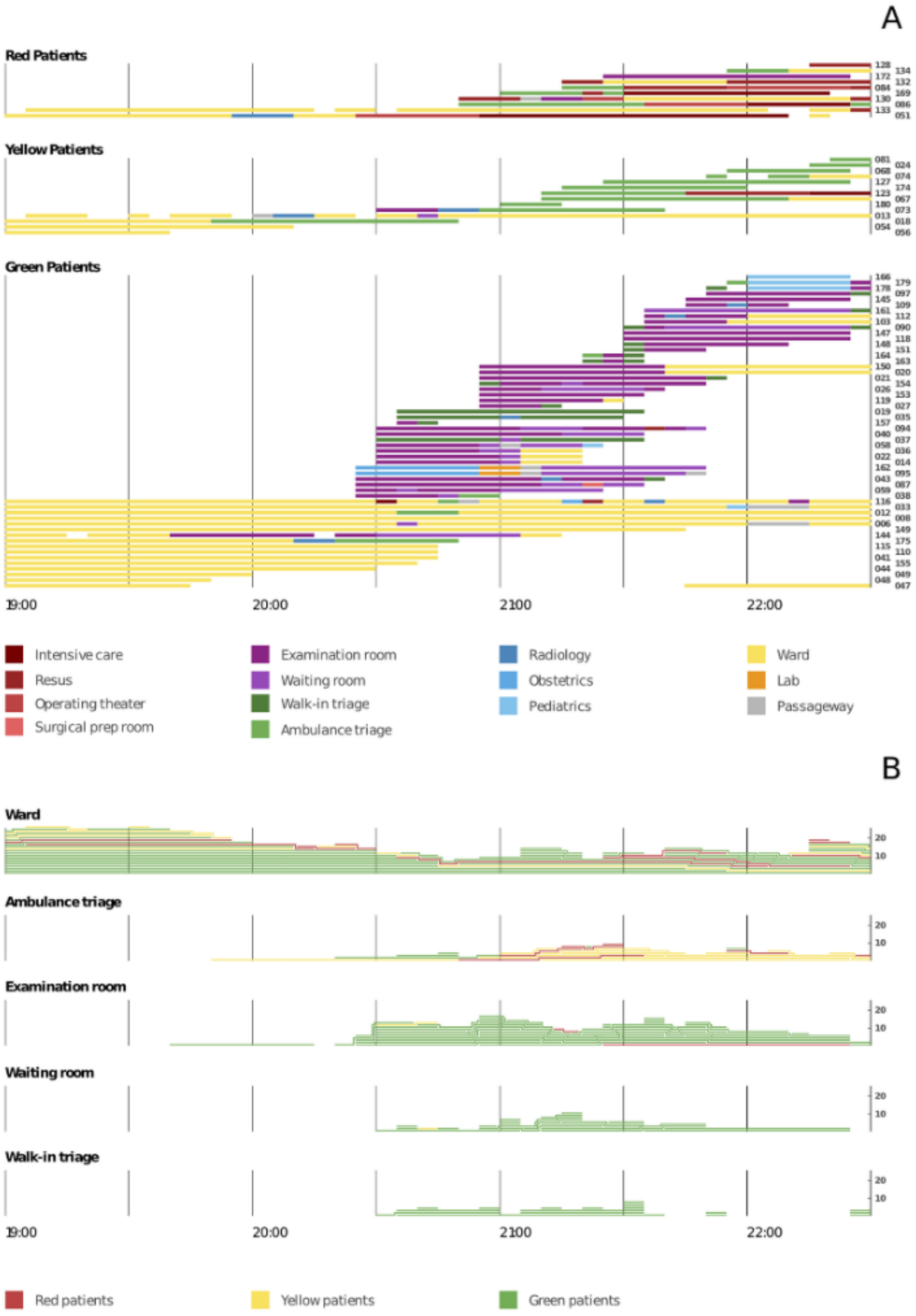
Figure 5. Cumulative time in contact (normalized on the total number of participants belonging to each caregiver category) between caregivers and victims with different triage at the Prehospital area (panel A) and Hospital area (panel B). EMS: Emergency Medical Services; HCA: Health Care Assistants; MD: Medical Doctors.



Casualty Flows in the Hospital

We studied the flow of 80 patients: 22 victims were already in the field hospital at the start of simulation as regular in-hospital patients and 58 victims were transferred from the accident site (the first transfer occurred at 8:15 pm). Figure 6, panel A, shows the victims flow through the field hospital of 56 patients with Green triage, 14 patients with Yellow, and 10 patients with Red triage. Each bar represents a patient, the color of the bar's segments refers to the room of the field hospital, and the length of the segments represents time passed by the victim in each room. The presence of patients in the different rooms of the hospital is consistent with triage code and diagnosis. Green victims passed most of their time in the Ward, Examination room, Walk-in triage room, and Waiting room. A total of 11 Yellow patients out of 14 spent time in the Ambulance triage, 6 Red victims out of 10 spent time in resus, and 3 Red victims spent time in the Intensive care. Figure 6, panel B, shows the number of victims in the Ward, Ambulance triage, Examination room, Waiting room, and Walk-in triage room over the simulation period. Each line corresponds to the presence of a victim; the color corresponds to the triage code.

Figure 6. Panel A: Victims flow through the field hospital. Each bar represents a patient (code is indicated on the right of the bar), the color of the bar's segments refers to the room of the field hospital, and the length of the segments represents time spent by the victim in each room. Numbers on the right part of the figures are the identification number for each victim. Panel B: Number of victims in the Ward, Ambulance triage, Examination room, Waiting room, and Walk-in triage over the simulation period. Each line corresponds to the presence of a victim; the color corresponds to the triage code. Numbers on the right part of the figure indicate the number of people in the room.



By analyzing the flow, we aimed at detecting the potential presence of bottlenecks in the field hospital. To do this, we focused on the analysis of the presence of the victims in the rooms in which they were not receiving any medical treatments. We defined bottlenecks as situations where the time of victims spent in the rooms where they did not receive any treatment is increased compared with the average time normally observed (for instance, when the number of victims in the hospital is low). Such rooms in the present settings are as follows: Ambulance triage, Examination room, Waiting room, and Walk-in triage room. We compared the numbers of victims present at the same time at the same room in relation with the waiting time in

Ambulance triage, Examination room, Waiting room, and Walk-in triage room; the Pearson correlation test was used. The mean waiting time in the Ambulance triage was 40 min (SD 9), in the Examination room was 40 min (SD 9), in the Waiting room was 39 min (SD 10), and in the Walk-in triage room was 31 min (SD 23). There were no significant correlations between the number of victims present in the same time at the same room and the waiting time; this result indicated that as the time passed in a room by a patient is not affected by the arrival of many victims in the hospital, in other terms, there was no obvious presence of bottlenecks.

We studied the presence patterns of individual victims in hospital rooms. For each patient we built a feature vector containing the time spent by that patient in each of the 15 hospital rooms normalized by total presence duration. The resulting set of 15-dimensional vectors (one per patient) was visualized using a dimensionality reduction technique known as t-SNE that maps each 15-dimensional patient vector to a 2-dimensional feature space (X_1 and X_2 axes; Figure 7). Clusters of patients with similar presence patterns are visible. We observed that the bottom-right cluster contains the more serious cases (yellow and red codes) and the top-right cluster contains the less serious cases (green codes), with the exception for the victim coded 172. The ideal triage code of the victim 172 was Yellow, and this victim passed the entire time of the simulation in the Examination room. On color coding by start location, the bottom-left cluster contains the majority of victims that started the simulation in the field hospital.

Moreover, we studied the presence times of individual victims in 3 hospital rooms characterized by longer presence times of the victims (Figure 8). We observed that the bottom-left group of patients on the t-SNE plot are characterized by long presence times in the Ward and they correspond to the victims that started the simulation in the field hospital; the group of victims on the top of the plot are characterized by long presence times in the Examination room, and they correspond to the victims coded green, and the patients that spent more time in the Ambulance triage room are the more serious cases, coded yellow and red.

Figure 7. Presence patterns of individual victims in hospital rooms visualized using t-Distributed Stochastic Neighbor Embedding (t-SNE) to map high-dimensional patient vectors to a 2-dimensional feature space (X_1 and X_2 axes). Each point corresponds to a patient. Victims with similar presence vectors are mapped to neighboring points in the plane. Victims are color coded by triage code (left panel) and by the start location of the simulation (red: accident site; blue: field hospital).

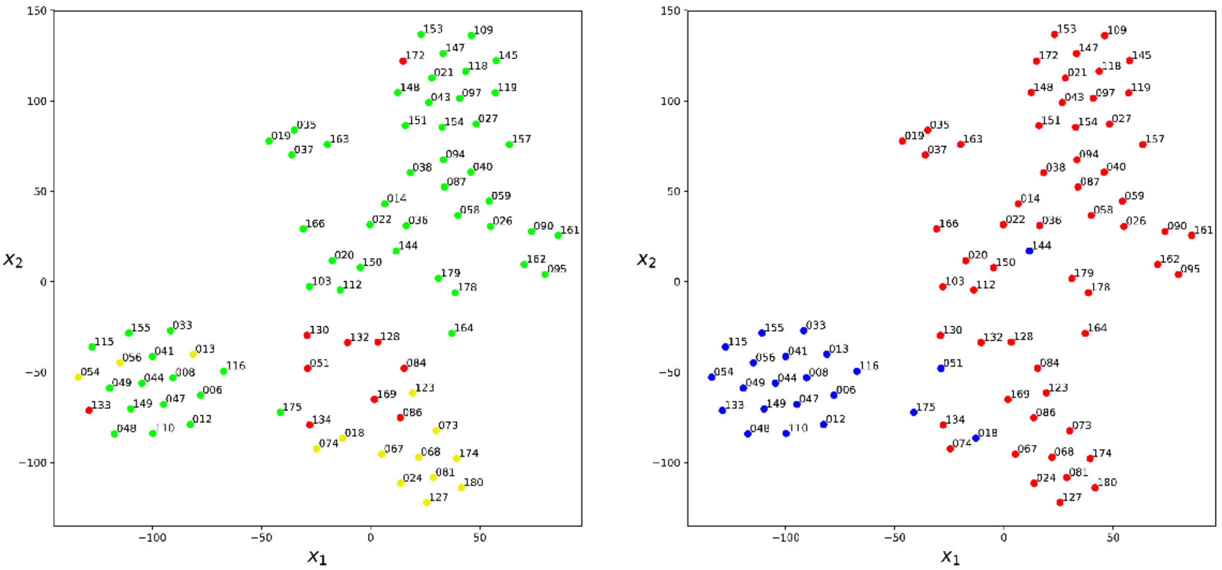
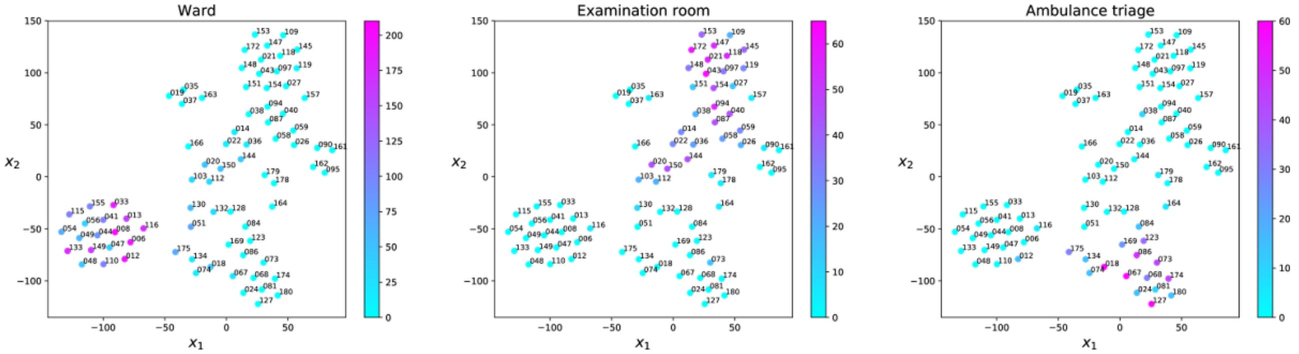


Figure 8. Presence patterns of individual victims in hospital rooms visualized using t-Distributed Stochastic Neighbor Embedding (t-SNE) to map high-dimensional patient vectors to a 2-dimensional feature space X_1 and X_2 axes). Each point corresponds to a patient. Victims with similar presence vectors are mapped to neighboring points in the plane. Points (patients) are color coded according to the time spent in minutes in the Ward (left panel), Examination room (middle panel), and Ambulance triage room (right panel).



Discussion

Principal Findings

With this study, we report the first quantitative assessment of social contact patterns in live MCI simulation, based on wearable proximity sensors. Our study showed the feasibility of the use of wearable proximity sensors to measure contact patterns during an MCI functional exercise. We obtained simple charts and contact matrices which allow for direct visualization of potentially missed opportunities for improvement of the MCI response.

In this study, we used proximity sensors to evaluate the contacts between individuals and the flow of the victims during an MCI simulation. The system provides information about the simulation which is coherent with the severity of the diagnosis both from the point of view of the relations between caregivers and victims as well as of the flow of victims. Our results showed that there were no differences between the median time passed by medical doctors and nurses with victims with different triage, both in the Prehospital and Hospital area. At the scene of the accident, this result is consistent with the chaotic environment typical of a disaster, where the medical staff task allocation is challenging and the medical interventions are equally distributed between patients with different severity injuries. However, significant differences between time in proximity by medical doctors and nurses with victims with different ISSs were observed at the Prehospital area. Both caregiver categories spent more time in contact with patients with the most severe injuries (classified as Critical) and, in particular, with patients classified as Moderate in terms of the ISS. When presented and discussed critically at the debriefing, it turned out that the majority of Moderate victims suffered bone fractures, and owing to this, the immobilization and stabilization procedures required a lot of time before being mobilized and transferred to the field hospital. At the Hospital area, nurses spent higher time in contact with Moderate and Serious patients in terms of the ISS.

Although there are no significant differences between the time passed by medical doctors and nurses with victims with different triage, proximity sensors revealed that medical doctors and nurses spent relevant time with patients with Red and Yellow triage, both in the Prehospital and Hospital area. Minor wounded (green codes) were predominantly managed by the EMS staff on scene and HCA in the hospital, allowing medical doctors and nurses to spend more time with the patients in most need. The quantitative measurement of contact patterns provided the opportunity to debate about it during the debriefing and identifying further strategies and counter actions. In disaster and MCI environments, coordination for task allocation is challenging. The analysis of temporal features of contact links between caregivers and casualties revealed proportional resource utilization of different health care skills for different victim severity triages and ISS codes.

Contact Data and Network Analysis

We obtained aggregated contact networks of the participants involved in the simulation, and we calculated the mean degree of the networks (ie, the mean number of connections between participants) for the whole duration of the simulation. Our results showed that the average degree was similar both for the in-Hospital and Prehospital area. The mean number of connections is lower for medical doctors and nurses than that of the EMS and HCA personnel. In other words, the medical staff interacted with a low number of patients by focusing treatment on a limited number of cases (see Section 3 of Multimedia Appendix 1). Moreover, the contact networks showed a high heterogeneity of the cumulative time spent in proximity (ie, weight of the edges) between participants, despite the short duration of the simulation. Our results show a highly heterogeneous distribution of contact durations characterized by a heavy tail; this outcome confirms the presence of a *universal* feature of contact patterns with most contacts of short duration and few long-lasting contacts. A similar general distribution of contact durations has been observed in other settings, including schools [20], hospitals [9], and households [22]. Moreover, the density of the networks (ie, fraction of all possible edges that are present in the network) was calculated to study the topology of networks built for each caregiver's categories and victims. The networks are sparse in both scenarios, in particular for EMS and HCA. However, we found that the density varied through the severity of injuries of the victims for the medical doctors and nurses, which showed a higher number of potential connections with victims with more serious conditions (see Section 4 in the Multimedia Appendix 1).

The analysis of temporal evolution of the number of contacts between participants revealed a high concentration of contacts during the middle part of the simulation at the Prehospital area, from 8 pm to 9:30 pm, even after the transfer of a part of the victims to the field hospital. The peak at the end of simulation is most likely due to an artefact: the meeting of participants shortly before the collections of sensors. In the Hospital area, the number of contacts was stationary until the transfer of patients from the Prehospital to Hospital area, then the number gradually increased as expected (see Section 5 in the Multimedia Appendix 1).

Flow and Presence Patterns of Victims in the Hospital

The deployment of sensors inside the hospital allowed to study the casualty flow. This analysis enabled to evaluate whether the patients were correctly headed by the health care personnel, consistently to the severity of the diagnosis and the expected location for such a diagnosis, in other words if patients with high acuity pathology were correctly occupying high acuity areas of the hospital and vice-versa, if low acuity patients were managed without wasting precious resources. Our results showed that the presence of patients in the hospital rooms were consistent with the triage code and diagnosis. Green victims spent most of their time in the examination room, Walk-in triage room, and Waiting room. A total of 6 Red victims out of 10 spent time in the Resus (ie, resuscitation area). A total of 3 Red patients spent most of their time in Intensive care (ie, Intensive Care Unit), and their diagnosis included head and chest trauma and septic shock. A total of 11 Yellow patients with minor injuries and stable trauma out of 14 spent time in the Ambulance triage. We studied the potential presence of bottlenecks for the rooms where victims were examined by medical staff, and they were waiting to receive treatments (Ambulance triage, Examination room, Waiting room, and Walk-in triage room) and evident presence of bottlenecks was not found. However, in our study, the number of victims in the field hospital is limited, and further evaluations of bottlenecks in simulations with a greater number of victims are necessary.

Similar presence patterns of victims coded with the same triage were observed in the rooms of the field hospital. In particular, the more serious cases (yellow and red codes) spent more time in the Ambulance triage room and the victims coded green spent more time in the Examination room. Our results showed a correspondence between the triage of the victims and the treatment given to the patients from the point of view of the permanence times in the rooms of the field hospital and the interactions with the caregivers (see Section 6 in the Multimedia Appendix 1). In other words, victims with comparable severity injuries were managed in a similar way in the field hospital. The red victim 172 was an exception in this trend: the ideal triage code of the victim 172 was yellow. This result is consistent with the high degree of overtriage of yellow casualties as red occurred in this simulation. It is well known that overtriage may cause fatigue of staff, depletion of resources, and impairment of efficient flow of critically injured patients through the system to definitive care [23]. The use of physical space plays a key role in managing a sudden influx of injured people or patients [24]. The evaluation of casualty flow and hospital space usage during exercises is a necessary first step in disaster preparedness and readiness by hospital authorities.

Limitations

It is important to highlight some limitations of this study. The exercise was organized as realistically as possible. Despite this, it is still a simulation, and the patients could portray only a limited number of changes in the clinical condition. Another potential issue concerns the possibility that participants changed their behavior because they were wearing sensors and knew they were participating in a scientific measure.

However, the methods presented in this paper can be useful to detect contact patterns in the very specific context of MCIs, thus allowing the implementation of tailored prevention strategies accordingly.

The measurement approach we used here also has limitations. Contacts were defined as face-to-face proximity, but no information on the possible occurrence of a physical contact between the 2 individuals is available [10], and consequently, no information on the interactions of caregiver-casualty is provided. Moreover, the short period of time of data collection (3 hours) also limits the ability to draw conclusions on what happens at longer time scales. However, through the use of the proximity-sensing platform, long-time studies are allowed.

Conclusions

In conclusion, our study showed that using wearable sensors based on proximity-sensor technology is feasible for obtaining a precise measurement of the pattern of close contacts among individuals during an MCI simulation. Although after-exercise debriefing sessions, during which participants discuss deficiencies warranting improvement, are routinely conducted, there is no commonly used and validated method for evaluating health performance during MCI exercises. Thus, our work constitutes a first step toward a standardized approach to the evaluation of an MCI exercise performance, as this monitoring system provides detailed temporal and spatial information about the medical staff and their interactions with the victims with limited human intervention. It represents, to our knowledge, the first example of unsupervised data collection of face-to-face contacts during an MCI exercise by means of wearable proximity sensors.

The unsupervised measurement of contact patterns with proximity sensors provides a unique opportunity to monitor the interactions between participants without the involvement of direct observers, which could impair the exercise's realism. Moreover, the analysis of contact patterns may help to identify specific interactions between health staff-patient to evaluate the decisions taken and the performance as the task allocation. In this study, the use of the sensors as fixed devices allowed to analyze the casualty flow in the field hospital to assess the use of physical space and resources allocation. The versatility of the system makes it possible to repeat similar studies in different environments, such as multiple vehicular accident settings or training for terrorist attacks, including smaller settings, and to compare results across contexts. Future studies could include a comparison of contact patterns on different settings of mass casualty simulations to improve the medical process, resource utilization, and decision making.

Acknowledgments

This study was supported by the Lagrange Project of the ISI Foundation funded by the CRT Foundation to LO, LG, MQ, MT, AP, KK, and CC. LC, PLI, DC, FDC, and CC designed the study. LO, LG, CC, and MQ analyzed the data. LO and LG drafted the manuscript. All authors critically revised and edited the draft and approved of the final version. LO, LG, and CC 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

LC, DC, and PLI received honoraria for lecturing and consulting in the field of disaster medicine training.

Multimedia Appendix 1

Additional description of the proximity sensing platform and additional analyses of the caregivers-victims network.

[PDF File (Adobe PDF File), 584KB - jmir_v21i4e12251_app1.pdf]

References

1. Admi H, Eilon Y, Hyams G, Utitz L. Management of mass casualty events: the Israeli experience. J Nurs Scholarsh 2011 Jun;43(2):211-219. [doi: 10.1111/j.1547-5069.2011.01390.x] [Medline: 21605326]
2. Zinan N, Puia D, Kinsley T. Results of a mass casualty incident simulation in an undergraduate nursing program. J Nurs Educ Pract 2015 Sep 22;5(12):71-78. [doi: 10.5430/jnep.v5n12p71]
3. Ingrassia PL, Carenzo L, Barra FL, Colombo D, Ragazzoni L, Tengattini M, et al. Data collection in a live mass casualty incident simulation: automated RFID technology versus manually recorded system. Eur J Emerg Med 2012 Feb;19(1):35-39. [doi: 10.1097/MEJ.0b013e328347a2c7] [Medline: 21659884]
4. Gillett B, Peckler B, Sinert R, Onkst C, Nabors S, Issley S, et al. Simulation in a disaster drill: comparison of high-fidelity simulators versus trained actors. Acad Emerg Med 2008 Nov;15(11):1144-1151 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00198.x] [Medline: 18717651]
5. Kaji AH, Langford V, Lewis RJ. Assessing hospital disaster preparedness: a comparison of an on-site survey, directly observed drill performance, and video analysis of teamwork. Ann Emerg Med 2008 Sep;52(3):195-201, 201.e1. [doi: 10.1016/j.annemergmed.2007.10.026] [Medline: 18191285]
6. Ingrassia PL, Prato F, Geddo A, Colombo D, Tengattini M, Calligaro S, et al. Evaluation of medical management during a mass casualty incident exercise: an objective assessment tool to enhance direct observation. J Emerg Med 2010 Nov;39(5):629-636. [doi: 10.1016/j.jemermed.2009.03.029] [Medline: 19570646]
7. Gao T, Pesto C, Selavo L, Chen Y, Ko J, Lim J, et al. Wireless medical sensor networks in emergency response: implementation and pilot results. 2008 Presented at: 2008 IEEE Conference on Technologies for Homeland Security; May 12-13, 2008; Waltham, MA, USA p. 187-192.
8. Jokela J, Rådestad M, Gryth D, Nilsson H, Rüter A, Svensson L, et al. Increased situation awareness in major incidents-radio frequency identification (RFID) technique: a promising tool. Prehosp Disaster Med 2012 Feb;27(1):81-87. [doi: 10.1017/S1049023X12000295] [Medline: 22575228]
9. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, van den Broeck W, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One 2011 Feb 28;6(2):e17144 [FREE Full text] [doi: 10.1371/journal.pone.0017144] [Medline: 21386902]
10. Vanhems P, Barrat A, Cattuto C, Pinton J, Khanafer N, Régis C, et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS One 2013 Sep;8(9):e73970 [FREE Full text] [doi: 10.1371/journal.pone.0073970] [Medline: 24040129]
11. Voirin N, Payet C, Barrat A, Cattuto C, Khanafer N, Régis C, et al. Combining high-resolution contact data with virological data to investigate influenza transmission in a tertiary care hospital. Infect Control Hosp Epidemiol 2015 Mar;36(3):254-260. [doi: 10.1017/ice.2014.53] [Medline: 25695165]
12. Della CF, Hubloue I, Ripoll GA, Ragazzoni L, Ingrassia PL, Debacker M. The European Masters Degree in Disaster Medicine (EMDM): a decade of exposure. Front Public Health 2014 May;2:49 [FREE Full text] [doi: 10.3389/fpubh.2014.00049] [Medline: 24904913]
13. Ingrassia P, Colombo D, Barra F, Carenzo L, Franc J, Della CF. Impact of training in medical disaster management: a pilot study using a new tool for live simulation. Emergencias 2013;25(459):459-466 [FREE Full text]
14. Carenzo L, Ragozzino F, Colombo D, Barra FL, Della CF, Ingrassia PL. Virtual Laboratory and Imaging: an online simulation tool to enhance hospital disaster preparedness training experience. Eur J Emerg Med 2018 Apr;25(2):128-133. [doi: 10.1097/MEJ.0000000000000421] [Medline: 27603200]
15. Benson M, Koenig KL, Schultz CH. Disaster triage: START, then SAVE--a new method of dynamic triage for victims of a catastrophic earthquake. Prehosp Disaster Med 1996;11(2):117-124. [doi: 10.1017/S1049023X0004276X] [Medline: 10159733]
16. Stevenson M, Segui-Gomez M, Lescohier I, di Scala C, McDonald-Smith G. An overview of the injury severity score and the new injury severity score. Inj Prev 2001 Mar;7(1):10-13 [FREE Full text] [doi: 10.1136/ip.7.1.10] [Medline: 11289527]
17. SocioPatterns. URL: http://www.sociopatterns.org/ [accessed 2018-08-02] [WebCite Cache ID 71N4LES0A]
18. Cattuto C, van den Broeck W, Barrat A, Colizza V, Pinton J, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS One 2010 Jul 15;5(7):e11596 [FREE Full text] [doi: 10.1371/journal.pone.0011596] [Medline: 20657651]
19. Stehlé J, Voirin N, Barrat A, Cattuto C, Colizza V, Isella L, et al. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. BMC Med 2011 Jul 19;9:87 [FREE Full text] [doi: 10.1186/1741-7015-9-87] [Medline: 21771290]
20. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J, et al. High-resolution measurements of face-to-face contact patterns in a primary school. PLoS One 2011 Aug;6(8):e23176 [FREE Full text] [doi: 10.1371/journal.pone.0023176] [Medline: 21858018]
21. Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579-2605 [FREE Full text]

22. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, et al. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. EPJ Data Sci 2016 Jun;5:21 [FREE Full text] [doi: 10.1140/epjds/s13688-016-0084-2] [Medline: 27471661]
23. Roccaforte JD, Cushman JG. Disaster preparedness, triage, and surge capacity for hospital definitive care areas: optimizing outcomes when demands exceed resources. Anesthesiol Clin 2007 Mar;25(1):161-77, xi. [doi: 10.1016/j.anclin.2007.01.002] [Medline: 17400163]
24. Sheikhbardsiri H, Raeisi AR, Nekoei-Moghadam M, Rezaei F. Surge capacity of hospitals in emergencies and disasters with a preparedness approach: a systematic review. Disaster Med Public Health Prep 2017 Dec;11(5):612-620. [doi: 10.1017/dmp.2016.178] [Medline: 28264731]

Abbreviations

DCC: dynamic casualty card
EMDM: European Master in Disaster Medicine
EMS: Emergency Medical Service
HCA: Health Care Assistant
ISS: Injury Severity Score
MCI: mass casualty incident
MD: Medical Doctors
RFID: Radio-Frequency IDentification
t-SNE: t-Stochastic Neighbor Embedding

Edited by G Eysenbach; submitted 19.09.18; peer-reviewed by O Leal Neto, S Erdley, E Prodromos; comments to author 16.01.19; revised version received 25.01.19; accepted 25.01.19; published 26.04.19.

Please cite as:
Ozella L, Gauvin L, Carenzo L, Quaggiotto M, Ingrassia PL, Tizzoni M, Panisson A, Colombo D, Sapienza A, Kalimeri K, Della Corte F, Cattuto C
Wearable Proximity Sensors for Monitoring a Mass Casualty Incident Exercise: Feasibility Study
J Med Internet Res 2019;21(4):e12251
URL: http://www.jmir.org/2019/4/e12251/
doi:10.2196/12251
PMID:31025944

Original Paper

Understanding User Experience: Exploring Participants' Messages With a Web-Based Behavioral Health Intervention for Adolescents With Chronic Pain

Annie T Chen[1], MSIS, PhD; Aarti Swaminathan[1], MS; William R Kearns[1], BS, BA; Nicole M Alberts[2], PhD; Emily F Law[3,4], PhD; Tonya M Palermo[3,4], MA, PhD

[1]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States
[2]Department of Psychology, St Jude Children's Research Hospital, Memphis, TN, United States
[3]Department of Anesthesiology and Pain Medicine, School of Medicine, University of Washington, Seattle, WA, United States
[4]Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, WA, United States

Corresponding Author:
Annie T Chen, MSIS, PhD
Department of Biomedical Informatics and Medical Education
School of Medicine
University of Washington
850 Republican Street, Box 358047, C238
UW Medicine South Lake Union
Seattle, WA, 98109
United States
Phone: 1 2062219218
Email: atchen@uw.edu

Abstract

Background: Delivery of behavioral health interventions on the internet offers many benefits, including accessibility, cost-effectiveness, convenience, and anonymity. In recent years, an increased number of internet interventions have been developed, targeting a range of conditions and behaviors, including depression, pain, anxiety, sleep disturbance, and eating disorders. Human support (coaching) is a common component of internet interventions that is intended to boost engagement; however, little is known about how participants interact with coaches and how this may relate to their experience with the intervention. By examining the data that participants produce during an intervention, we can characterize their interaction patterns and refine treatments to address different needs.

Objective: In this study, we employed text mining and visual analytics techniques to analyze messages exchanged between coaches and participants in an internet-delivered pain management intervention for adolescents with chronic pain and their parents.

Methods: We explored the main themes in coaches' and participants' messages using an automated textual analysis method, topic modeling. We then clustered participants' messages to identify subgroups of participants with similar engagement patterns.

Results: First, we performed topic modeling on coaches' messages. The themes in coaches' messages fell into 3 categories: Treatment Content, Administrative and Technical, and Rapport Building. Next, we employed topic modeling to identify topics from participants' message histories. Similar to the coaches' topics, these were subsumed under 3 high-level categories: Health Management and Treatment Content, Questions and Concerns, and Activities and Interests. Finally, the cluster analysis identified 4 clusters, each with a distinguishing characteristic: Assignment-Focused, Short Message Histories, Pain-Focused, and Activity-Focused. The name of each cluster exemplifies the main engagement patterns of that cluster.

Conclusions: In this secondary data analysis, we demonstrated how automated text analysis techniques could be used to identify messages of interest, such as questions and concerns from users. In addition, we demonstrated how cluster analysis could be used to identify subgroups of individuals who share communication and engagement patterns, and in turn facilitate personalization of interventions for different subgroups of patients. This work makes 2 key methodological contributions. First, this study is innovative in its use of topic modeling to provide a rich characterization of the textual content produced by coaches and participants in an internet-delivered behavioral health intervention. Second, to our knowledge, this is the first example of the use of a visual analysis method to cluster participants and identify similar patterns of behavior based on intervention message content.

(J Med Internet Res 2019;21(4):e11756) doi:10.2196/11756

KEYWORDS

data visualization; natural language processing; chronic pain; cluster analysis; technology

Introduction

Background

In recent years, an increased number of internet interventions have been developed in behavioral health, targeting a range of conditions and behaviors, including depression, pain, anxiety, substance use, sleep disturbance, psychotic disorders, and eating disorders. Delivery of interventions on the internet offers many benefits, including accessibility, cost-effectiveness, convenience, and anonymity [1,2]. However, though research has shown that internet behavioral interventions work, questions remain concerning how, why, and for whom, and there remains a need for better methods for investigating these questions [3-5]. There are gaps in our knowledge, including the need to identify predictors of therapeutic success or failure [6], reasons for attrition and dropout [7], long-term impacts, active intervention components, and methods for tailoring and promoting engagement [8-11]. Predicting which patients will respond best to which treatments [12] and personalizing interventions [13] are also important challenges. Finally, demonstrating treatment fidelity by showing that a treatment is applied consistently to all participants randomized to a treatment is critical [14].

Fortunately, with internet-delivered interventions, we are also seeing a diverse array of passively collected data which can help us better understand participants' experiences. In this study, we are particularly concerned with textual data (eg, messages exchanged with a coach), which could, for example, provide insight into participants' experiences and enable us to tailor interventions to their needs. Though it has been observed that there is potential to employ natural language processing techniques to personalize mental health treatments, to date, such examples are scarce [15].

Textual data could also help us learn more about the effects of coaching and intervention components. A few studies of adult populations have used content analysis to categorize and explore associations between coach behaviors (eg, task reinforcement and self-efficacy shaping) and treatment response [16-18]. Content analysis of client emails has also shown that attempts to try alternative behaviors and observation of positive consequences are associated with engagement, as measured by module completion [19,20]. Thus, developing better methods to examine textual data is important for a number of reasons, including personalization, assessment of the effectiveness of coaching, and understanding whether participants are learning and practicing the treatment content.

However, content analysis can be time-consuming and laborious. Some studies have used automatic methods of textual analysis, such as the Linguistic Inquiry and Word Count (LIWC; Pennebaker Conglomerates, Inc) software, to analyze communication patterns in therapeutic settings and online support groups [21-26]. The LIWC assesses word usage in particular domains such as positive and negative emotion, anxiety, and pronouns [27], but it does not facilitate thematic analysis. To better characterize the subject matter of communications, we could consider other automated forms of textual analysis, such as topic modeling [28] and document clustering [29,30], which have been used to evaluate health communication patterns in social networking platforms and online health communities, but not textual data from internet-delivered interventions.

Objectives

In this study, we aimed to demonstrate the feasibility of a novel textual and visual analytic approach to identify patterns of engagement during an internet-delivered cognitive behavioral therapy (CBT) intervention for youth with chronic pain aged 11 to 17 years and their parents [31]. We have previously demonstrated the efficacy of an internet CBT intervention versus internet-delivered pain education on our primary outcome of adolescent activity limitations at 6-month follow-up [31]. In this secondary analysis, we focus only on youth randomized to the internet CBT arm because this was the only treatment arm in our trial that included human support. In the internet CBT intervention, participants could interact with a coach via an asynchronous message center. First, we present findings from topic modeling to characterize the primary themes of the messages sent by coaches and participants. The purpose of this step was to present an overview of the thematic content as well as to demonstrate how the results of topic modeling can differ depending on the content author. Second, we focus on intervention participants, employing cluster analysis to identify and visualize subgroups of participants with similar patterns of treatment engagement and message content.

Our work is novel in 2 ways: (1) in the use of topic modeling to provide a rich characterization of intervention participants' experiences and (2) through the development of a visual method for comparing textual artifacts of participant engagement. We take a visual analytics approach, which facilitates interpretation of complex data by combining concepts from data mining, machine learning, human-computer interaction, and human cognition [32]. In health care, we often see it applied to electronic health records for the analysis of patient trajectories and to identify patients with similar clinical characteristics (eg, [33,34]). Extant literature has also included visual analysis of health-related communication [28,35,36], but to our knowledge, there has not been prior work using visual methods to examine coach-participant messages in internet interventions.

Methods

Internet-Delivered Cognitive Behavioral Therapy

Adolescents and parents in the internet CBT condition received access to Web-based Management for Adolescent Pain (Web-MAP2), a pain self-management intervention based on cognitive-behavioral, social learning, and family systems theories. Treatment content and program features have been described in detail elsewhere [31] (see Figure 1 for a screenshot of the home page). The program has a travel theme with each treatment module representing a different destination from around the world. Adolescents and parents were provided with access to separate, password-protected websites that included treatment modules, audio files of relaxation exercises, a progress tracker, and a message center where they could exchange asynchronous messages with a coach.

Adolescent treatment modules included the following: (1) pain education, (2) recognizing stress and negative emotions, (3) relaxation methods, (4) coping with pain at school, (5) cognitive coping skills, (6) sleep hygiene and lifestyle skills, (7) increasing activity, and (8) relapse prevention. Parent treatment modules included the following: (1) pain education, (2) recognizing stress and negative emotions, (3) operant strategies I (attention and praise), (4) operant strategies II (reward systems and strategies to support school goals), (5) modeling, (6) sleep hygiene and lifestyle, (7) communication skills, and (8) relapse prevention. Adolescents and parents were asked to complete 1 treatment module per week, which was designed to be analogous to weekly sessions delivered in face-to-face CBT. In 6 of the 8 modules, adolescents and parents were asked to practice coping skills and complete weekly behavioral assignments related to their personalized goals.

In this secondary data analysis, we focused on understanding the content of messages exchanged between coaches and participants (adolescents and parents) in the message center. Through the message center, coaches provided participants with personalized feedback about each behavioral assignment. Adolescents and parents could also initiate messages to the coach via the message center at any time. During the trial, all assignments were reviewed by 1 of the 5 coaches with prior experience in CBT (4 psychology postdoctoral fellows and 1 masters-level therapist). Coaches responded to each behavioral assignment and all messages initiated by participants. Coaches used a manual to standardize all messages sent to participants, which emphasized rapport building (eg, "What do you like to do for fun?"), praise for skills practice (eg, "Great job spending more time in school this week!"), and strategies to overcome barriers to skills practice (eg, "Try practicing deep breathing at the same time every day."). Coaches were supervised by a licensed clinical psychologist (TP) via regular message review to ensure adherence to the manual and standardization of messages sent to participants.

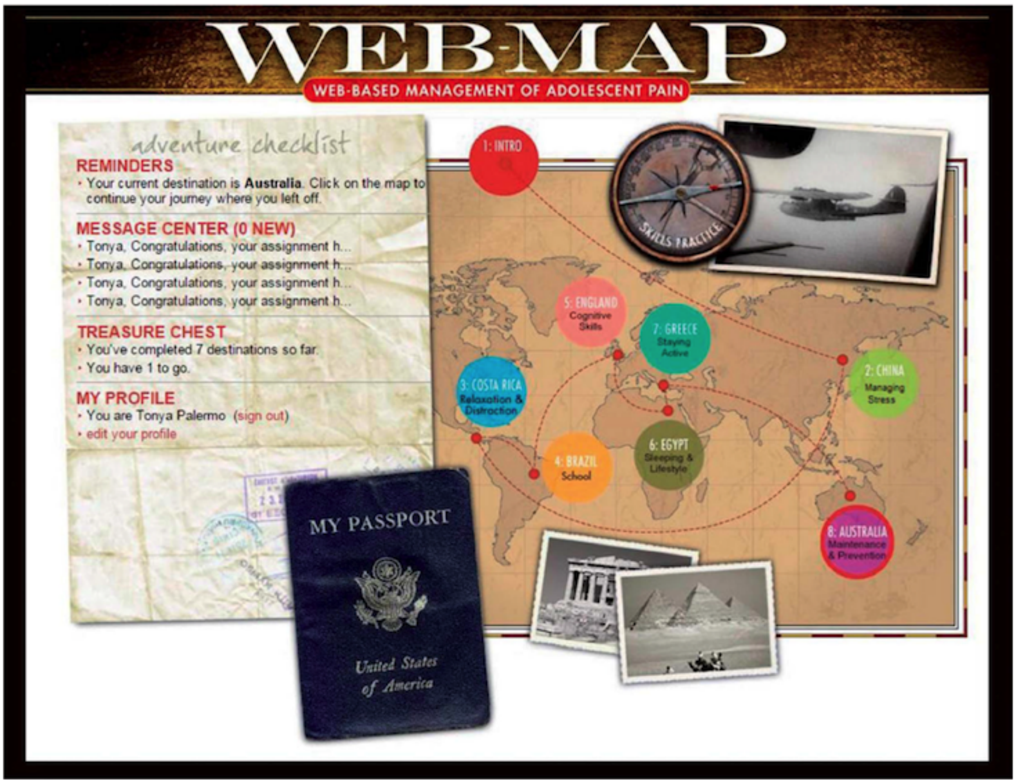
The study was approved by the primary site's institutional review board and the institutional review boards at each referring center. Adolescents gave assent and parents provided informed consent before initiating any research procedures.

Sample and Measures

The data for this study are from the internet-delivered CBT arm (n=138) of the trial (N=273) [31]. The inclusion criteria were as follows: (1) adolescent aged between 11 and 17 years, (2) adolescent reporting chronic idiopathic pain present over the previous 3 months, (3) adolescent reporting pain at least once per week, (4) parent reporting pain interfering with at least one area of daily functioning, and (5) the adolescent received a new patient evaluation in 1 of the participating pain clinics.

As this study focused on the messages sent by the coaches, teens, and parents, we excluded the participants who did not send any messages (n=15). In addition, 1 participant was excluded after the topic modeling and cluster analysis because they did not meet the eligibility criteria for the main study. We report the statistics for this sample (n=123).

Figure 1. Web-based Management for Adolescent Pain (Web-MAP) homepage.



REMINDERS
- Your current destination is **Australia**. Click on the map to continue your journey where you left off.

MESSAGE CENTER (0 NEW)
- Tonya, Congratulations, your assignment h...
- Tonya, Congratulations, your assignment h...
- Tonya, Congratulations, your assignment h...
- Tonya, Congratulations, your assignment h...

TREASURE CHEST
- You've completed 7 destinations so far.
- You have 1 to go.

MY PROFILE
- You are Tonya Palermo (sign out)
- edit your profile

At pretreatment, parents reported on adolescents' age, gender, and racial background via the secure study website. Similar to other studies, we also measured engagement using unobtrusive measures of participant exposure and skills practice [11,37,38]. Specifically, we employed 4 measures of engagement: the number of messages sent, the average word count of participants' messages, module completion rates, and percentage of interactive fields.

Interactive fields in all the modules facilitated personalization of the intervention. For example, in the module focused on parent operant training, parents were asked to input the behaviors they wanted to target in a behavior management plan; the behaviors that parents entered were subsequently used to remind parents what they want to focus on in the next week [39]. As completion of the interactive fields was not required, the percentage of interactive fields completed could be considered an indicator of the level of engagement.

Using Topic Modeling to Characterize the Message Content

To identify topics discussed by coaches and participants, we employed a generative probabilistic modeling algorithm, Latent Dirichlet Allocation (LDA), to identify the main themes in the messages. LDA models documents as random mixtures over topics, where a topic is defined as a distribution of words [40]. The output of LDA is the distribution of topics within each document and a word distribution for each topic [41]. The keywords for each topic can then be used, along with documents that are predicted to have high proportions of the topics, to determine what the topics are. However, examining the documents along with their predicted probabilities for each topic is not necessarily an intuitive way to explore a document collection. Thus, in recent years, we have seen various efforts to create visual ways of exploring the topics (eg, [41-43]). In this section, we explain the details of the topic modeling procedure that we performed, and later, we will explain the visual application that we developed for exploring the textual data for internet-delivered interventions.

We used the LDA implementation available within the MAchine Learning for LanguagE Toolkit (MALLET) toolkit [44] to identify the most common topic within each message. We experimented with varying numbers of topics ranging from 10 to 45 and elected to use a 15-topic solution, which provided a balance between diversity of topics and ease of interpretability. Using a greater number of topics could lead to greater precision in terms of topic but could also make it more difficult for researchers to interpret.

When performing topic modeling, it is helpful to exclude words such as "a," "an," and "the" that appear frequently but add little meaning. To accomplish this, we employed the default stop word list that is provided with the MALLET toolkit, along with a custom stop word list consisting primarily of first names, to exclude coaches' and participants' names. We performed topic modeling separately on coaches' and participants' messages because there were marked differences in these messages, and combining them decreased topic coherence.

We verified topic assignments by randomly selecting 20 messages categorized under each topic and manually checking to see if the topic assignment was accurate. Where there were less than 20 messages pertaining to a topic, all messages assigned to that topic were verified. For participants, the accuracy rate was 76.5% (192/251). For coaches, the accuracy rate was 99.3% (298/300).

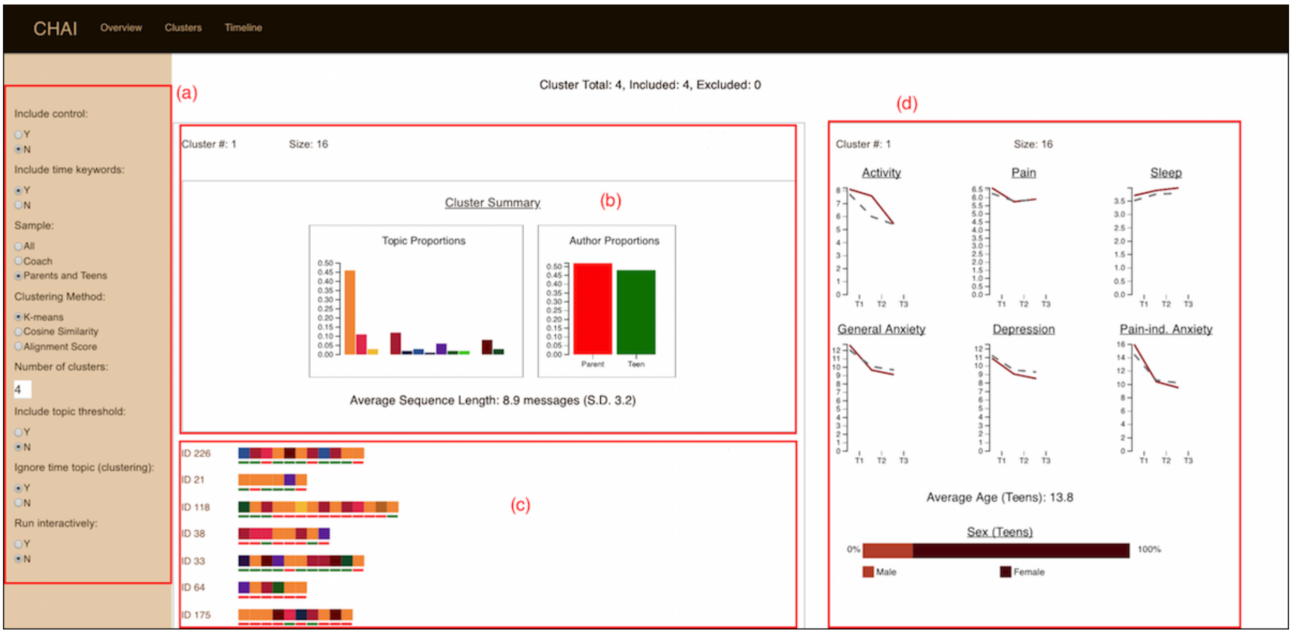
Clustering and Visual Analysis of Participants' Message Histories

We used cluster analysis to identify subgroups of participants that shared similarities in their message histories. Cluster analysis was performed using the Communication History Analysis Interface (CHAI), a visual interface that we developed which offers users the capability to visualize participants' message histories, perform cluster analysis, and explore the results of cluster analysis.

To identify subgroups of participants with similar message histories, we employed the k-means clustering method [45] to cluster parent and teen pairs by the topics that they discussed with their coaches. K-means cluster analysis takes a set of n-dimensional points and clusters them into a set of K clusters [45]. Each parent and teen pair's communications with the coaches were represented using a 15-dimensional vector, 1 dimension for each topic identified in the topic modeling procedure. To give an example, suppose a parent and teen pair authored 6 messages in total, 2 each for topics 3, 5, and 7. Their contribution would be represented by {0, 0, 2, 0, 2, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0}. Thus, the vector representing each parent and teen pair would illustrate common topics within that pair's communications, and the results of the cluster analysis would yield parent and teen pairs that discussed similar topics. As 1 topic, Time, was highly prevalent and yet had no specific meaning other than the references to time, it was excluded from the clustering.

We employed 2 methods together, visual examination and the inverse scree plot [46], to select the number of clusters. We plotted the variance for solutions with the number of clusters k varying from 1 to 20, and selected 4 as the optimal solution for 2 reasons. At this point the increasing the number of clusters led to less substantial decreases in variance, but there was not a "clear bend." We visually examined solutions of differing numbers through the CHAI interface, deciding on 4 to err on the side of coarser clusters that illustrated differences in participants' textual communications, but did not differentiate too granularly within the sample. The k-means clustering method can be susceptible to the starting seeds [47]. To avoid bias, we repeated the clustering with different starting seeds and observed that the defining characteristics of the clustering solutions remained the same in the repetitions.

Figure 2. Communication history analysis interface: (a) cluster controls, (b) cluster engagement characteristics (theme proportions and parent/teen participation), (c) message sequences, and (d) cluster demographics.



Include control:
○ Y
● N
Include time keywords:
● Y
○ N
Sample:
○ All
○ Coach
● Parents and Teens
Clustering Method:
● K-means
○ Cosine Similarity
○ Alignment Score
Number of clusters:
4
Include topic threshold:
○ Y
● N
Ignore time topic (clustering):
● Y
○ N
Run interactively:
○ Y
● N

Cluster Total: 4, Included: 4, Excluded: 0

Cluster #: 1 Size: 16

Average Sequence Length: 8.9 messages (S.D. 3.2)

Average Age (Teens): 13.8

We examined the results of the cluster analysis using the CHAI application that we developed. The clustering feature of this application features 2 primary views, an Overview of the clustering results (Figure 2) and a Cluster Detail view that can be used to examine the messages for each cluster. CHAI provides summaries of cluster engagement characteristics that show the prevalence of all topics in each cluster, so that users can compare the clusters in terms of topic and authorship. The CHAI application performs cluster analysis and displays participants' message histories by cluster. For any given participant identification (ID) number, each message history is rendered as a horizontal sequence, with the earliest message to the left and the last message on the right. The right pane enables users to view outcomes and demographic characteristics for each cluster. The CHAI application was developed using Python, the machine learning library scikit-learn, and Web development frameworks and Javascript visualization libraries including AngularJS and D3.

Characterizing the Clusters

To compare clusters, analysis of variance (ANOVA) is often used for interval/ratio variables and Chi-square analyses for categorical variables [48-50]. In this study, we employed a hybrid comparison method involving both visual analysis and statistical measures. A visual analysis of the message histories and engagement summaries for each cluster enabled us to characterize each cluster's communication patterns in terms of topical emphasis. We then employed statistical measures to compare the clusters further. We performed ANOVA to compare the clusters in terms of the number of messages, message length, module completion rates, and percentage of interactive fields completed. We also compared the clusters on demographic characteristics using ANOVA for teen age and Fisher exact test for teen gender because of expected cell counts of less than 5 [51].

Results

Sample and Measures

Descriptive statistics for adolescents are presented in Table 1. The majority of the sample was female and white, with a mean age of 14.7 (SD 1.6) years. Overall, both parents and teens were highly engaged (Table 2). Descriptive statistics for the message data are presented in Table 3.

Table 1. Teen demographics (N=123).

Demographic characteristic	Statistics
Age (years), mean (SD)	14.7 (1.6)
Teen gender, n (%)	
Male	26 (21.1)
Female	97 (78.9)
Teen racial background, n (%)	
White	115 (94.3)
African American	2 (1.6)
Native Hawaiian or other Pacific Islander	1 (0.8)
American Indian or Alaskan Native	1 (0.8)
Mixed	3 (2.5)

Table 2. Engagement measures.

Measure	Child, mean (SD)	Parent, mean (SD)
Modules completed	7.5 (1.3)	7.3 (1.5)
Percentage of interactive fields	79.6 (15.4)	72.8 (19.6)

Table 3. Message data.

Message type	Messages, n	Word count, mean (SD)
All messages	3426	132.7 (91.1)
By coaches	2692	149.8 (87.7)
By parents	347	88.6 (87.0)
By teens	387	52.7 (55.1)

Using Topic Modeling to Characterize the Message Content

Primary Themes in the Coaches’ Message Content

We performed topic modeling and identified 15 topics. The themes in coaches’ messages fell into 3 categories: Treatment Content, Administrative and Technical, and Rapport Building (Table 4; Multimedia Appendix 1). These categories are consistent with those that we employed in our prior work [39].

As expected, as the responses followed a coaching manual, there was a great deal of consistency in the topics and their order. In Figure 3, we see examples of coaches’ message histories, with each series corresponding to a different parent and teen. Certain topics appeared almost universally, such as the coach’s initial greeting in the beginning and the instructions to complete the Web-based diary at the end. Summation of progress and encouragement occurred toward the end of the intervention. Other topics such as relaxation skills, working toward goals, and lifestyle changes reflected the treatment content and followed the order of modules in Web-MAP2. This visualization can help us to verify that treatment is being delivered consistently to intervention participants.

Treatment Content

The main themes of the coaches’ messages reflected the treatment content. Some topics, such as Lifestyle Changes, complemented the skills that participants learned. Other topics, including guidance on how to use Web-MAP2 and touching base on participants’ progress, were not tied to particular treatment content:

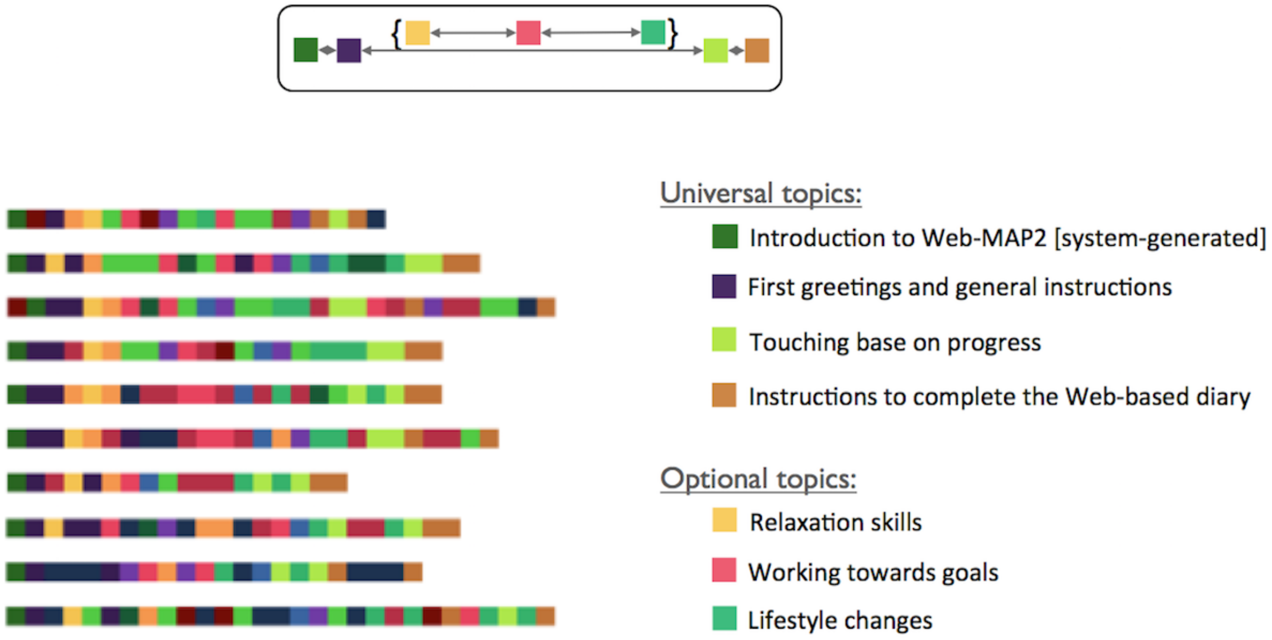
> *It sounds like you're trying to change your eating habits which is a great idea. It can be hard to eat when you don't really feel hungry! Adding snacks throughout the day of food that used to be enjoyable to you is one strategy to help kick-start your appetite.* [ID 190, in Lifestyle changes]
>
> *I’ve looked at the progress tracker in your Passport and [participant name]’s pain sleep and ability to do things with friends have steadily improved since your family started Web-MAP2. That’s great! Congratulations to you both!* [ID 286, in Touching base on progress]

Table 4. Topics in coaches' messages.

Topic number	Main theme	Keywords	Percentage[a]
Treatment Content			
1	Reinforcing behaviors in parents	Great destination praise assignment congratulations move behaviors praising approved sounds job positive nice work staff negative	4.6
2	Relaxation skills	Relaxation breathing practice practicing time deep feel bit find destination relaxing congratulations assignment skills approved good mp move audio	3.8
3	Working toward goals	School plan work system great assignment goal destination reward approved move congratulations staff working sounds goals set time relaxation	8.9
6	Encouraging parents to share their coping strategies	Strategies stress great coping helpful life stressed job approved part move types learn normal destination assignment congratulations staff work	4.3
8	Thought replacement, thought stopping and relaxation techniques	Practice relaxation skills work great job day home practicing thoughts good thought exercises helpful assignment minutes mp ipod school	6.9
10	Encouragement and strategies of how to utilize the program	Skills pain program work good learning staff hear great logging helpful find question week encourage manage sounds strategies time	7.4
12	Lifestyle changes	Great sleep work lifestyle goal water habits making sounds goals set good congratulations assignment move approved destination make specific	8.7
13	Touching base on progress	Program skills pain learned ability great passport participate activities progress continue tracker refresher return congratulations finished job looked anytime	8.2
Administrative and Technical			
5	Instructions/reminders to complete the Web-based diary	Home staff question online section complete reminders back follow diary destination show directions visited link passed shown months pointing	10.3
7	Responding to questions and/or information about assignments	Assignment program week complete staff questions message assessment destination days time touch log working send destinations approve back completed	7.5
9	First greeting to participants and general instructions	Week destination assignment questions log complete clues program stress learning strategies approved working move find end things pain coach	11.2
11	Introduction to Web-MAP2 and general instructions	Introduction journey home destination online complete final destinations questions begin staff start participation path proceed baby awhile camps parades	5.4
Rapport Building			
4	Responding to participants' descriptions of activities, interests, and family	Great fun sounds time nice work weekend staff logging family hear good friends awesome play week things busy message	8.0
14	Expressing empathy, followed by constructive feedback	Time family teens feel pain good system important make start day behavior reward goals problem choose challenging things	3.2
15	Asking for updates about life and general treatment progress	Note hear things drop update move assignment congratulations staff destination time approved great fun sounds messaging working system job	1.6

[a] Proportion of messages that were assigned to this topic.

Figure 3. Temporal patterns in coaches' messages. Web-MAP: Web-based Management for Adolescent Pain.



Administrative and Technical

This category included the generic, fully scripted reminders or introductions that the system sent, the administrative messages that coaches sent (eg, self-introductions and reminders), and coaches' response to administrative and technical questions:

> *Now that you've visited all of your Web-MAP2 destinations we'd like you to again complete your online diary.* [ID 249, in Instructions/reminders to complete the Web-based diary]
>
> *It might be helpful to read the Brazil readings today/tomorrow so you can think about the assignment over the week and submit the assignment this weekend.* [ID 209, in Responses to questions and/or information about assignments]

Rapport Building

Some topics focused on building rapport with participants, empathizing with them and providing support and guidance for incorporating the skills participants learned into their own lives. The position of these topics within a message sequence was unrelated to participants' progress through the modules:

> *Thanks for sharing a little bit more about your family [participant name] it's great that you're all so active and enjoy sports. I'm really glad to hear that [participant name] has found a way to continue to be involved through coaching--when she is ready to play again it will make the transition a lot easier* [ID 222, in Responses to participants' descriptions of activities, interests, and family]

Primary Themes in the Participants' Message Content

We employed topic modeling to participants' messages. Overall, 15 topics were identified, and similar to the coaches' topics, these appeared to be subsumed under 3 high-level categories: Health Management and Treatment Content, Questions and Concerns, and Activities and Interests (Table 5; Multimedia Appendix 2). In addition, 2 topics, Family and Time, were reflections of common linguistic usage but not necessarily thematically important and will not be reviewed in detail in this section.

Health Management and Treatment Content

One of the main themes in participants' messages was their health management and treatment content. Many of the messages illustrated how participants were incorporating the skills that they were learning in their own lives:

> *Its been a long road but finally [participant name] is feeling better. Her pain is now a 6 out of 10 when it was a 10 for the longest time. It definately helps to have the education and training that you have provided* [ID 53, in Progress in learning pain and stress management techniques]
>
> *hello: one way that we have changed is to let one another finish what is being said before jumping in: another way is the tone that is used: its a more calmer tone rather than a frustrated tone.* [ID 15, in Rewards system, coping and achieving goals]

Questions and Concerns

The Questions and Concerns category included 3 topics. Questions and Concerns about Assignments included updates that participants had completed assignments as well as inquiries about technical problems that they had regarding assignment completion, such as the system not storing that the assignment had been completed and asking participants to do the assignment again. The Suggestions topic included messages in which participants asked for guidance for a problem or made a suggestion about Web-MAP2. Finally, the Questions topic included a diverse range of questions, many regarding gift cards or technical errors:

> *I just wanted to let you know that I did answer the questions in this assignment at the beginning of the week. I wanted to reread some notes yesterday and when I went backwards in the lesson it must have reset my answers from earlier in the week.* [ID 215, in References to assignments]
>
> *I was wondering if you have any suggestions for ways that I could stay healthy during summer when my routine is not regular.* [ID 74, in Suggestions]

Table 5. Topics in participants' messages.

Topic number	Themes	Keywords	Percentage
Health Management and Treatment Content			
3	Progress in learning pain and stress management techniques	Pain time great program things good talk helpful learning techniques life support skills starting working level stress study dealing	11.8
4	Pain	Pain school time day back work hard week days bad past program feel working put weeks started made problem	16.1
5	Medications, nutrients and lab results	Prescribed day attention mg caffeine vitamin continued pill results symptoms November needed relief normal work bleeding progesterone bcp uti	0.5
11	Rewards system, coping and achieving goals	System reward kids rewards privilege points worked motivated shows exams weekly sessions studying encouraged totally sign pretty talked fighting	3.3
14	Fatigue, sleep, relaxation techniques	Sleep relaxation bed find breathing staying sleeping imagery helping practice asleep fatigue deep muscle hours helps night extremely told	4.6
Questions and Concerns			
1	References to assignments	Back week complete assignment completed destination message brazil finished wanted log finish point logged diary destinations end costa thought	13.6
2	Suggestions	Suggestions tq interested making things email walking process change putting small positive concerns important visit healthy medical challenge strategies	2.2
8	Questions	Question make class teacher card place gift work continue found amazon email principal music special wondering privileges comfortable pm	3.8
Activities and Interests			
15	Fun with family and friends	fun time friends play enjoy family day meet things nice great watch playing games soccer read pretty kids favorite	16.6
6	Creative arts	Dog back show book mom walk picture sit dress pretty ups light continue black room hit front middle lab	3.3
7	Music, sports, and school	Year years grade team school high people drive place plays called chorus trombone band telling funny students graduation glad	1.5
10	Drama and reading	Drama schools rock club theatre reading show festival play England kohl hear bit excited players shows production piano props	2.4
13	Trips	Home trip water softball july track crew cheer short till checking change fine feet vomiting sounds fruit taking grandma	1.9
Other Topics			
9	Family	Year family husband children close home brothers call therapy lives told seattle blood older years thinks needed turned occasionally	2.2
12	Time	Week school weekend good home back year today mom start happy wait work family assignment coming thing end break	16.6

Activities and Interests

A significant portion of participants' messages described activities, interests and hobbies they enjoyed. The topics were primarily differentiated according to particular hobbies and interests, as specified by their corresponding keywords. These topics demonstrated ways in which their pain and other aspects of their health might interfere with their activities and how participants reacted:

> *I am into sports and other activity's and when i am doing something I will start hurting but I don't do anything about it because I will not let the pain stop me from doing what I am doing.* [ID 104, in Fun with family and friends]

Clustering and Visual Analysis of Participants' Message Histories

We performed k-means clustering and selected 4 as the optimal clustering solution as described in the Methods section. Each of the clusters had a distinctive characteristic, either in terms of topic or extent of communication (Figure 4). Statistics for cluster membership and participation appear in Table 6.

Characterizing the Clusters

Assignment-Focused

The defining characteristic of this cluster was the prevalence of messages relating to assignments (Figure 4, orange). Participants were diligent about completing tasks and giving updates whenever they were delayed. Salient recurrent issues included technical problems and confusion about what they were supposed to be doing. With regard to the former, participants often reported that they had completed an assignment, but were asked to do the assignment again, leading to confusion and frustration. With regard to the latter, participants were sometimes generally confused about the program, but there were also more specific causes of confusion, for example, when lessons called for doing something related to school and school was not in session. Overall, there was a significant difference in age between the clusters ($F_{3,119}$=3.1; P=.03). As the average age of the Assignment-Focused cluster was lower than the other groups, they perhaps needed more direction than the intervention participants in other clusters.

Short Message Histories

This cluster was the largest of the 4 and included approximately half of the sample. The distinguishing characteristic of this cluster was that there were significantly fewer messages sent as compared with the other clusters (Table 6), and there were no striking patterns in the topics discussed. Participants' messages tended to be responses to questions from the coaches or to assignments. There were some technical questions, but otherwise, participants rarely reached out themselves to start a conversation or overtly ask for guidance. There were a fair number of apologetic utterances by participants, explaining that they had not spent time on Web-MAP2 because of other commitments.

Pain-Focused

In this cluster, participants' messages focused primarily on pain and secondarily on pain management and activities (Figure 4, dark, light pink, and brown, respectively). Both parents and teens described the pain and other health issues that they experienced and the efforts that they made to deal with these issues. Their health status had a clear impact on their lives, in terms of their schedules and daily routines. Health care systems and health care providers also played a prominent role in conversations. Some of the messages in this cluster were long, with participants presenting detail concerning issues that they were having, reductions in their pain, or telling coaches about strategies that had had a positive effect.

Activity-Focused

A substantial part of the message content of this cluster was related to activities. These messages tended to have a more conversational feel, with participants sharing excitement and other sentiments toward the activities that they were engaged in, such as going to concerts, participating in drama, reading, and camping. There was also content concerning participants' integration of the skills that they were learning. There were occasional questions of a relationship building nature, in which participants shared what they were doing, and then asked coaches questions about them, for example, what they had done over the weekend and what they liked to do. Like the Pain-Focused cluster, teens tended to engage with the coaches more than the parents.

Comparing Engagement Patterns Across Clusters

Aside from the number of messages, we investigated whether the clusters differed in terms of engagement through 3 other measures: message word counts, module completion rates, and percentage of interactive fields completed. The average message word count was not significantly different across clusters. For both the other types of engagement, the differences were significant at the .05 level for teens but not for parents (Table 6). Overall, the patterns across clusters were similar to what we observed with the number of messages. The parents and teens in the Short Message Histories cluster had the lowest module completion rates and interactive fields completed, with 1 exception: the parents in the Assignment-Focused cluster exhibited the lowest percentage of fields completed.

Figure 4. Topic proportions, parent-teen message proportions, and representative topics for each cluster. Red underline denotes a parent-authored message, and green underline denotes a teen-authored message.

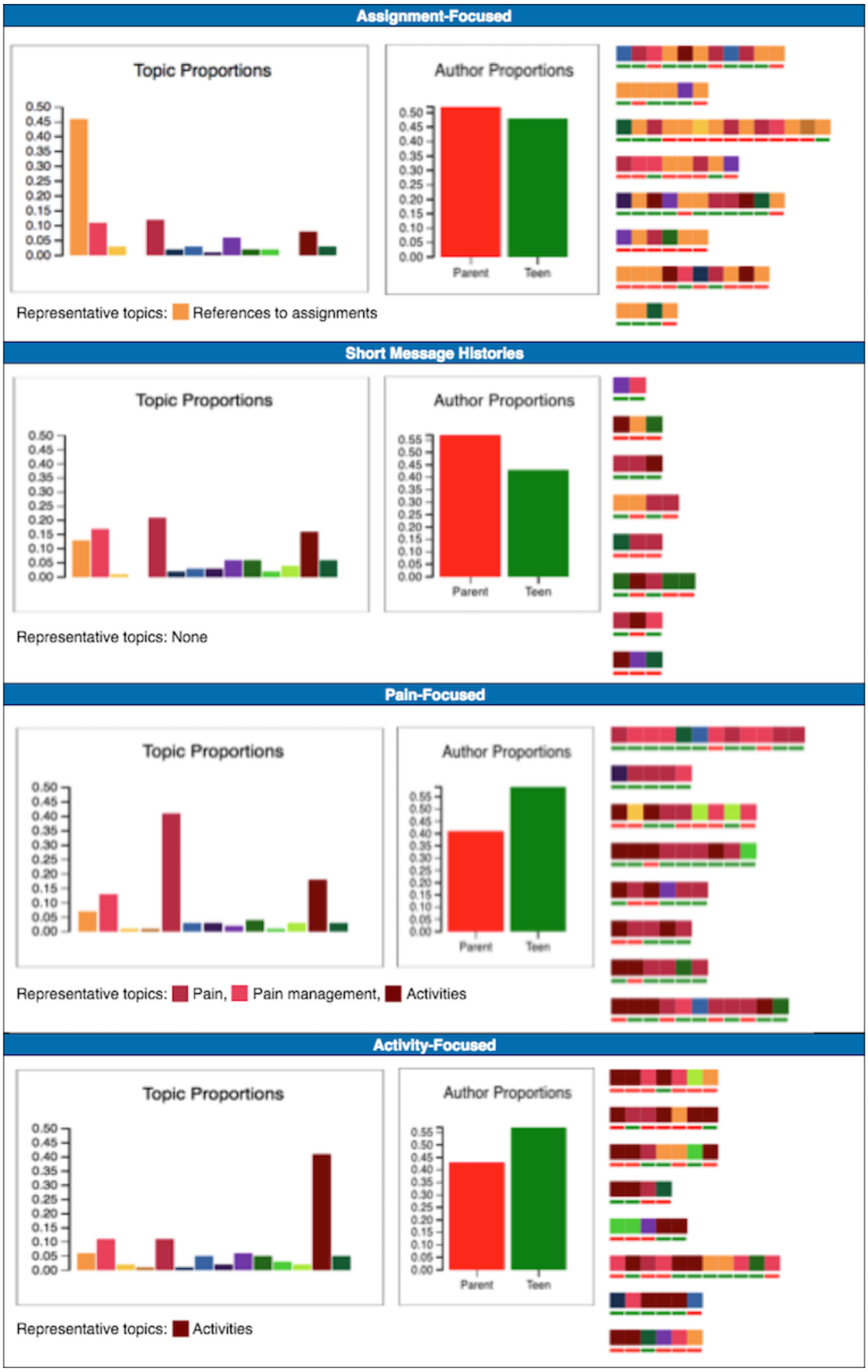


Table 6. Cluster membership and participation.

Comparison dimension	Assignment-Focused (n=16)	Short Message Histories (n=62)	Pain-Focused (n=20)	Activity-Focused (n=25)	*F* value (df)	*P* value
Age (years), mean (SD)						
Teen	13.80 (1.4)[a]	15.01 (1.6)[b]	14.69 (1.7)	14.27 (1.6)	3.1 (3,119)	.03[c]
Gender, observed/expected						.93[d]
Male	3/3	14/13	3/4	6/5	—[e]	—
Female	13/13	48/49	17/16	19/20	—	—
Number of messages, mean (SD)						
Teen	4.38 (3.5)[a]	1.65 (1.7)[b,f,g]	5.4 (2.8)[a]	4.24 (3.0)[a]	16.2 (3,119)	<.001[c]
Parent	4.56 (3.0)[a]	1.95 (1.6)[b,f]	3.80 (2.7)[a]	3.04 (2.0)	8.7 (3,119)	<.001[c]
Message word count, mean (SD)						
Teen	32.23 (1.0)	25.00 (1.9)	32.28 (1.0)	34.90 (0.4)	2.1 (3,119)	.10
Parent	41.7 (1.0)	42.58 (1.9)	46.39 (1.1)	46.64 (0.4)	0.3 (3,119)	.80
Module completion rate, mean (SD)						
Teen	7.63 (0.9)	7.13 (1.6)[g]	7.65 (0.8)	7.92 (0.4)[a]	2.8 (3,117)	.05[c]
Parent	7.44 (1.0)	7.00 (1.9)	7.6 (1.1)	7.88 (0.4)	2.5 (3,117)	.06
Interactive fields completed, mean (SD)						
Teen	82.73 (12.0)	75.18 (18.8)[g]	83.07 (8.5)	85.58 (7.4)[a]	3.8 (3,119)	.01[c]
Parent	68.84 (17.3)	70.27 (22.5)	72.72 (19.5)	81.81 (7.6)	2.4 (3,119)	.07

[a] The given cluster is significantly different from the Short Message Histories cluster at the .05 level.

[b] The given cluster is significantly different from the Assignment-Focused cluster at the .05 level.

[c] Significant at the .05 level.

[d] Fisher's Exact Test was performed because the gender variable is categorical and had expected cell values of less than 5 (please refer back to the Methods section for more detail).

[e] Not applicable.

[f] The given cluster is significantly different from the Pain-Focused cluster at the .05 level.

[g] The given cluster is significantly different from the Activity-Focused cluster at the .05 level.

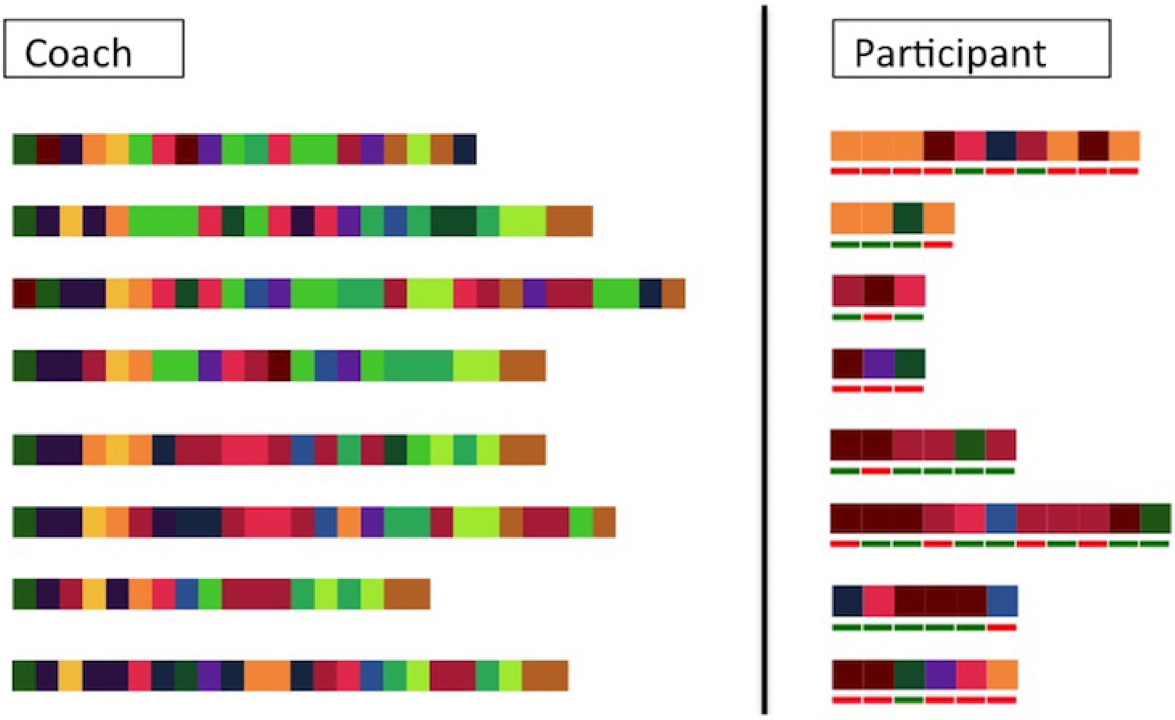
Discussion

Principal Findings

In this secondary data analysis, we aimed to demonstrate the feasibility of employing a text and visual analytics approach to automatically characterize the intervention experience and identify subgroups of users with similar participation patterns in an internet-delivered behavioral health intervention. Our approach had 2 main parts. First, we employed automated text analysis methods to identify the primary themes of the messages sent by coaches as well as adolescent and parent users. Using a technique called topic modeling, we identified 15 topics from the coaches' messages, which were subsumed under the high-level categories of Treatment Content, Administrative and Technical, and Rapport Building. We also examined the main themes that were discussed by adolescents and their parents, identifying 15 topics subsumed under 3 high-level categories: Health Management and Treatment Content, Questions and Concerns, and Activities.

Second, we employed k-means cluster analysis to identify subgroups of participants who shared similarities with one another in terms of their message content. The solution included 4 clusters: Assignment-Focused, Short Message Histories, Pain-Focused, and Activity-Focused. As compared with the other clusters, the Assignment-Focused cluster experienced proportionately greater issues and confusion with the program. Participants in the Short Message Histories cluster engaged significantly less with the coaches than the other clusters. The Pain-Focused and Activity-Focused clusters shared more with the coaches, with the thematic foci indicated in the cluster names.

Figure 5. Temporality and the lack of temporality in coach versus participant sequences. Colored blocks denote messages and their assigned topics. Please see Multimedia Appendices 1 and 2 for color keys.



These clusters reflect different tendencies in the topics of conversation and interaction patterns within the sample. If cluster profiles could be formed in real time, participants' interaction patterns and preferences could be considered in personalizing interventions. For example, if we are able to detect early on that certain participants need more feedback about assignments, coaches can offer additional guidance or the system can provide additional scaffolding for those participants. Whereas the coaches' messages exhibited a clear temporal progression, the main message characteristics of the participant clusters demonstrated no such progression, suggesting that without additional guidance, engagement characteristics shown by participants are likely to remain consistent over time (Figure 5). If additional guidance or scaffolding were provided to these participants, we could then observe whether the thematic content of participants' messages changes after the introduction of this additional support.

Methodological Implications

We now consider the methodological contributions of the work presented. First, we demonstrated that topic modeling can produce coherent themes in textual data from internet-delivered interventions. The themes that were reported in the coaches' messages were consistent with the topics covered in the Web-MAP2 modules, and the emergent topics reflected therapist skills, such as task reinforcement and encouragement, that have been reported in previous research on therapists' behaviors [16-18]. This concordance suggests that topic modeling could potentially be used to assess treatment fidelity in internet-delivered behavioral health interventions as an alternative to qualitative content analysis methods that may be more time-consuming and as a way to categorize data in a format that could be used in subsequent systematic analyses, such as the cluster analysis described in this study.

However, despite the promise that has been shown in the use of topic modeling as a technique for facilitating the analysis of textual data from internet-delivered interventions, the study also identified areas for improvement. The accuracy of the topic modeling algorithm on the participants' messages was not as high as on the coaches' messages. There are perhaps several reasons for this. First, as the coaches were expected to deliver the same treatment to all participants, the consistency of the messages led to better performance of the algorithm. In addition, the number of coaches' messages was substantially greater than the number of participants' messages. There is a need to consider ways to improve the performance of the topic modeling of patient messages. One possibility might be to employ a method that incorporates domain knowledge about what types of topics that we expect, such as seeded LDA [52].

A second major contribution of this work is the development of a visual method for depicting and comparing sequential textual data from intervention participants. The visual representation that we employed in this study facilitates quick identification of temporal characteristics of message data as well as comparison of message sequences. Moreover, the auxiliary visualizations (topic proportion and author proportion) that we developed facilitate characterization of cluster members' engagement with the intervention through their textual contributions. This is a significant contribution to the body of research concerning visualization of temporal health care data, which has often focused on visualizing structured data from electronic health records [53], though there are a few examples of visualization of conversational data collected in clinical settings (eg, [36,54]).

The work that we present here also suggests possibilities for just-in-time monitoring of behavioral health interventions. In the case of online forums, coaches or moderators often have a high volume of messages to monitor, and automated detection of messages of interest can reduce this burden [55]. Previous work has employed natural language processing methods to identify messages suggesting recovery problems in a substance abuse forum [55], and threads in an online diabetes community requiring moderator assistance [56].

Our results suggest that there is potential for the use of the techniques developed here to flag messages and issues for coaches to follow up on. In this study, the Questions and Concerns category was perhaps the one of greatest interest, as it included messages in which participants could benefit most from additional guidance. Automatically identifying these types of messages can provide insight into additional areas in which participants need information or support. Though preliminary, these results suggest that a dashboard could be developed that hides routine messages, identifies messages of interest, and categorizes and organizes issues for coaches to address. To increase the viability of such a dashboard, there is a need for additional work, involving input from intervention coaches, to ensure that the visual displays are clinically relevant and meaningful.

Limitations and Future Directions

Our analysis has various limitations. First, we identified groups of individuals with shared characteristics in terms of content and volume of communication, measured through the number of messages sent. There may be a need for richer characterization of participant experiences for the purposes of tailoring and personalization. To do so, one might consider examining clustering solutions with a larger number of clusters or employing additional features in the cluster analysis to represent other dimensions of participant experience.

Second, in this study, we focused on the participants in our cluster analysis. In the future, we plan to develop visual methods to examine the dyadic interaction between coaches and participants. If we are able to identify frequent patterns of interaction and their consequences, then this information could help us to better understand how to provide support and guidance to participants during the course of an intervention. In our cluster analysis, we did not consider the timing of the topics in participants’ trajectories. In future work, it could be helpful to combine the temporality of messages, as well as other types of participant actions, in cluster analysis. Finally, we did not examine whether the patterns of interaction relate to treatment outcomes; future research is needed to understand the potential impact of interaction patterns on treatment benefit from internet-delivered interventions.

Conclusions

In this study, we combined text and visual analytics techniques to explore messages authored in an internet-delivered behavioral health intervention for adolescents with chronic pain and their parents. We employed topic modeling to identify the main topics discussed by coaches and participants. Doing so helped us to characterize coaches’ behaviors and important aspects of participants’ experiences. Using cluster analysis and visual analytics, we identified participants who shared similarities in the ways that they interacted with coaches during the intervention. To our knowledge, this is the first example of the use of a visual analysis method employing textual data collected from an internet-delivered behavioral health intervention to cluster participants and identify similar patterns of behavior. Taking the entirety of participants’ engagement patterns—their topics of discussion, information needs, and interaction patterns into consideration—could potentially facilitate personalization and tailoring of interventions.

Acknowledgments

This study was supported by NIH R01HD062538 (PI: TMP), and we would like to thank the children and families who participated. WRK’s work on this study was supported by NIH National Library of Medicine Biomedical and Health Informatics Training Grant at the University of Washington Grant Nr. T15LM007442, and EFL’s work on this study was supported by NIH K23NS089966.

Authors' Contributions

ATC conceptualized the study and wrote the majority of the text; AS contributed major sections to the paper; ATC, AS, and WRK developed the visual interface that is described in the paper; ATC and AS performed the data analyses described; and NMA, EFL, and TMP provided feedback on the data analysis and on the manuscript as it was being revised. All authors have provided approval for the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Topics in coaches’ messages.

[PDF File (Adobe PDF File), 129KB - jmir_v21i4e11756_app1.pdf]

Multimedia Appendix 2

Topics in participants' messages.

[PDF File (Adobe PDF File), 76KB - jmir_v21i4e11756_app2.pdf]

References

1. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Jul;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
2. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
3. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2011 May 7 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada p. 3063-3072 URL: http://dl.acm.org/citation.cfm?id=1979396 [doi: 10.1145/1978942.1979396]
4. Schueller SM, Munoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 3;22(6):478-483. [doi: 10.1177/0963721413495872]
5. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes internet therapy work? Cogn Behav Ther 2009 Jan;38 Suppl 1:55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
6. 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]
7. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
8. Cunningham JA, Gulliver A, Farrer L, Bennett K, Carron-Arthur B. Internet interventions for mental health and addictions: current findings and future directions. Curr Psychiatry Rep 2014 Dec;16(12):1-5. [doi: 10.1007/s11920-014-0521-5] [Medline: 25308390]
9. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
10. Short CE, Rebar AR, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Eur Health Psychol 2015;17(1):32-38 [FREE Full text]
11. 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 Dec;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
12. Bagby RM, Quilty LC, Segal ZV, McBride CC, Kennedy SH, Costa JP. Personality and differential treatment response in major depression: a randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy. Can J Psychiatry 2008 Jun;53(6):361-370 [FREE Full text] [doi: 10.1177/070674370805300605] [Medline: 18616856]
13. Ng MY, Weisz JR. Annual Research Review: building a science of personalized intervention for youth mental health. J Child Psychol Psychiatry 2016 Mar;57(3):216-236 [FREE Full text] [doi: 10.1111/jcpp.12470] [Medline: 26467325]
14. Resnick B, Bellg AJ, Borrelli B, Defrancesco C, Breger R, Hecht J, et al. Examples of implementation and evaluation of treatment fidelity in the BCC studies: where we are and where we need to go. Ann Behav Med 2005 Apr;29(Suppl):46-54. [doi: 10.1207/s15324796abm2902s_8] [Medline: 15921489]
15. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Sep;23(5):649-685. [doi: 10.1017/S1351324916000383]
16. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]
17. Holländare F, Gustafsson SA, Berglind M, Grape F, Carlbring P, Andersson G, et al. Therapist behaviours in internet-based cognitive behaviour therapy (ICBT) for depressive symptoms. Internet Interv 2016 Mar;3:1-7. [doi: 10.1016/j.invent.2015.11.002]
18. Schneider LH, Hadjistavropoulos HD, Faller YN. Internet-delivered cognitive behaviour therapy for depressive symptoms: an exploratory examination of therapist behaviours and their relationship to outcome and therapeutic alliance. Behav Cogn Psychother 2016 Nov;44(6):625-639. [doi: 10.1017/S1352465816000254] [Medline: 27302220]
19. Svartvatten N, Segerlund M, Dennhag I, Andersson G, Carlbring P. A content analysis of client e-mails in guided internet-based cognitive behavior therapy for depression. Internet Interv 2015 May;2(2):121-127. [doi: 10.1016/j.invent.2015.02.004]

20. Soucy JN, Hadjistavropoulos HD, Couture CA, Owens VA, Dear BF, Titov N. Content of client emails in internet-delivered cognitive behaviour therapy: a comparison between two trials and relationship to client outcome. Internet Interv 2018 Mar;11:53-59. [doi: 10.1016/j.invent.2018.01.006]
21. Dirkse D, Hadjistavropoulos HD, Hesser H, Barak A. Linguistic analysis of communication in therapist-assisted internet-delivered cognitive behavior therapy for generalized anxiety disorder. Cogn Behav Ther 2015 Sep;44(1):21-32. [doi: 10.1080/16506073.2014.952773] [Medline: 25244051]
22. Owen JE, Yarbrough EJ, Vaga A, Tucker DC. Investigation of the effects of gender and preparation on quality of communication in Internet support groups. Comput Hum Behav 2003 May;19(3):259-275. [doi: 10.1016/S0747-5632(02)00068-7]
23. van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
24. Siriaraya P, Tang C, Ang CS, Pfeil U, Zaphiris P. A comparison of empathic communication pattern for teenagers and older people in online support communities. Behav Inf Technol 2011 Sep;30(5):617-628. [doi: 10.1080/0144929X.2011.582146]
25. Lieberman MA, Goldstein BA. Not all negative emotions are equal: the role of emotional expression in online support groups for women with breast cancer. Psychooncology 2006 Feb;15(2):160-168. [doi: 10.1002/pon.932] [Medline: 15880627]
26. Kramer AD, Fussell SR, Setlock LD. Text analysis as a tool for analyzing conversation in online support groups. In: CHI '04 Extended Abstracts on Human Factors in Computing Systems. New York, NY: ACM; 2004 Apr 24 Presented at: CHI EA'04; April 24-29, 2004; Vienna, Austria p. 1485-1488 URL: http://dl.acm.org/citation.cfm?id=986096 [doi: 10.1145/985921.986096]
27. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The University of Texas at Austin. 2015. The development and psychometric properties of LIWC2015 URL: https://repositories.lib.utexas.edu/handle/2152/31333 [accessed 2019-03-03] [WebCite Cache ID 76bQjEKJi]
28. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015 Sep;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]
29. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013 Feb;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]
30. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]
31. 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]
32. Caban JJ, Gotz D. Visual analytics in healthcare--opportunities and research challenges. J Am Med Inform Assoc 2015 Mar;22(2):260-262. [doi: 10.1093/jamia/ocv006] [Medline: 25814539]
33. Wang TD, Wongsuphasawat K, Plaisant C, Shneiderman B. Extracting insights from electronic health records: case studies, a visual analytics process model, and design recommendations. J Med Syst 2011 Oct;35(5):1135-1152. [doi: 10.1007/s10916-011-9718-x] [Medline: 21541691]
34. Gotz D, Sun J, Cao N, Ebadollahi S. Visual cluster analysis in support of clinical decision intelligence. AMIA Annu Symp Proc 2011;2011:481-490 [FREE Full text] [Medline: 22195102]
35. MacLean D, Hangal S. You Didn't Tell Me That! Visualizing the Hidden Attributes of Online Health Communities. 2009 Oct 4 Presented at: ACM Symposium on User Interface Software and Technology (UIST); October 4-7, 2009; Victoria, BC, Canada p. 1-4.
36. Angus D, Watson B, Smith A, Gallois C, Wiles J. Visualising conversation structure across time: insights into effective doctor-patient consultations. PLoS One 2012 Jun;7(6):e38014 [FREE Full text] [doi: 10.1371/journal.pone.0038014] [Medline: 22693629]
37. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
38. Doherty G, Coyle D, Sharry J. Engagement with Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2012 May 5 Presented at: CHI'12; May 5-10, 2012; Austin, TX p. 1421-1430. [doi: 10.1145/2207676.2208602]
39. Law EF, Murphy LK, Palermo TM. Evaluating treatment participation in an internet-based behavioral intervention for pediatric chronic pain. J Pediatr Psychol 2012 Sep;37(8):893-903 [FREE Full text] [doi: 10.1093/jpepsy/jss057] [Medline: 22511033]
40. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022.
41. Ganesan A, Brantley K, Pan S, Chen J. arXiv. 2015 Jul 23. LDAExplore: Visualizing Topic Models Generated Using Latent Dirichlet Allocation URL: https://arxiv.org/abs/1507.06593 [accessed 2019-03-03] [WebCite Cache ID 76bRjFKUR]

42. Chaney AJ, Blei DM. Visualizing Topic Models. : Association for the Advancement of Artificial Intelligence; 2012 Jun 4 Presented at: Sixth International AAAI Conference on Weblogs and Social Media (ICWSM); June 4-7, 2012; Dublin, Ireland p. 1-4 URL: http://www.cs.columbia.edu/~blei/papers/ChaneyBlei2012.pdf
43. Chuang J, Manning CD, Heer J. Termite: visualization techniques for assessing textual topic models. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. New York, NY: ACM; 2012 May 21 Presented at: AVI'12; May 21-25, 2012; Capri Island, Italy p. 74-77 URL: http://dl.acm.org/citation.cfm?id=2254572 [doi: 10.1145/2254556.2254572]
44. McCallum AK. MAchine Learning for LanguagE Toolkit. 2002. MALLET: A Machine Learning for Language Toolkit URL: http://mallet.cs.umass.edu/ [accessed 2019-03-03] [WebCite Cache ID 76bSKGYpz]
45. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]
46. Lathrop RG, Williams JE. The reliability of inverse scree tests for cluster analysis. Educ Psychol Meas 1987;47(4):953-959. [doi: 10.1177/0013164487474010]
47. Khan SS, Ahmad A. Cluster center initialization algorithm for K-means clustering. Pattern Recognit Lett 2004 Aug;25(11):1293-1302. [doi: 10.1016/j.patrec.2004.04.007] [Medline: 25904163]
48. Yukselturk E, Top E. Exploring the link among entry characteristics, participation behaviors and course outcomes of online learners: an examination of learner profile using cluster analysis. Br J Educ Technol 2012 Jul 9;44(5):716-728. [doi: 10.1111/j.1467-8535.2012.01339.x] [Medline: 24843434]
49. Ares G, Gámbaro A. Influence of gender, age and motives underlying food choice on perceived healthiness and willingness to try functional foods. Appetite 2007 Jul;49(1):148-158. [doi: 10.1016/j.appet.2007.01.006] [Medline: 17335938]
50. Fotopoulos C, Krystallis A, Vassallo M, Pagiaslis A. Food Choice Questionnaire (FCQ) revisited. Suggestions for the development of an enhanced general food motivation model. Appetite 2009 Feb;52(1):199-208. [doi: 10.1016/j.appet.2008.09.014] [Medline: 18929606]
51. McDonald JH. Handbook of Biological Statistics. Baltimore, MD: Sparky House Publishing; 2009.
52. Ramesh A, Goldwasser D, Huang B, Daume H, Getoor L. Understanding MOOC discussion forums using seeded LDA. 2014 Jun 26 Presented at: Ninth Workshop on Innovative Use of NLP for Building Educational Applications; June 26, 2014; Baltimore, MD, USA p. 28-33.
53. Rind A, Wang TD, Aigner W, Miksch S, Wongsuphasawat K, Plaisant CC, et al. Interactive information visualization to explore and query electronic health records. Found Trend Hum–Comput Interact 2011;5(3):207-298. [doi: 10.1561/1100000039]
54. Bartels J, Rodenbach R, Ciesinski K, Gramling R, Fiscella K, Epstein R. Eloquent silences: a musical and lexical analysis of conversation between oncologists and their patients. Patient Educ Couns 2016 Dec;99(10):1584-1594 [FREE Full text] [doi: 10.1016/j.pec.2016.04.009] [Medline: 27156659]
55. Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, et al. Detecting recovery problems just in time: application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. J Med Internet Res 2018 Jun 12;20(6):e10136 [FREE Full text] [doi: 10.2196/10136] [Medline: 29895517]
56. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013 Dec;46(6):998-1005 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.011] [Medline: 24025513]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
CHAI: Communication History Analysis Interface
LDA: Latent Dirichlet Allocation
LIWC: Linguistic Inquiry and Word Count
Web-MAP2: Web-based Management for Adolescent Pain

Edited by G Eysenbach; submitted 31.07.18; peer-reviewed by C Ang; comments to author 14.10.18; revised version received 05.02.19; accepted 10.02.19; published 15.04.19.

Please cite as:
Chen AT, Swaminathan A, Kearns WR, Alberts NM, Law EF, Palermo TM
Understanding User Experience: Exploring Participants' Messages With a Web-Based Behavioral Health Intervention for Adolescents With Chronic Pain
J Med Internet Res 2019;21(4):e11756
URL: http://www.jmir.org/2019/4/e11756/
doi:10.2196/11756
PMID:30985288

Original Paper

Wearable Sensors Reveal Menses-Driven Changes in Physiology and Enable Prediction of the Fertile Window: Observational Study

Brianna Mae Goodale[1*], PhD; Mohaned Shilaih[1*], PhD; Lisa Falco[1], PhD; Franziska Dammeier[1], PhD; Györgyi Hamvas[1], MSc; Brigitte Leeners[2]

[1]Ava AG, Zurich, Switzerland

[2]Department of Reproductive Endocrinology, University Hospital, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Brianna Mae Goodale, PhD
Ava AG
Gutstrasse 73
Zurich, 8055
Switzerland
Phone: 41 788737163
Email: brianna.goodale@avawomen.com

Abstract

Background: Previous research examining physiological changes across the menstrual cycle has considered biological responses to shifting hormones in isolation. Clinical studies, for example, have shown that women's nightly basal body temperature increases from 0.28 to 0.56 ˚C following postovulation progesterone production. Women's resting pulse rate, respiratory rate, and heart rate variability (HRV) are similarly elevated in the luteal phase, whereas skin perfusion decreases significantly following the fertile window's closing. Past research probed only 1 or 2 of these physiological features in a given study, requiring participants to come to a laboratory or hospital clinic multiple times throughout their cycle. Although initially designed for recreational purposes, wearable technology could enable more ambulatory studies of physiological changes across the menstrual cycle. Early research suggests that wearables can detect phase-based shifts in pulse rate and wrist skin temperature (WST). To date, previous work has studied these features separately, with the ability of wearables to accurately pinpoint the fertile window using multiple physiological parameters simultaneously yet unknown.

Objective: In this study, we probed what phase-based differences a wearable bracelet could detect in users' WST, heart rate, HRV, respiratory rate, and skin perfusion. Drawing on insight from artificial intelligence and machine learning, we then sought to develop an algorithm that could identify the fertile window in real time.

Methods: We conducted a prospective longitudinal study, recruiting 237 conception-seeking Swiss women. Participants wore the Ava bracelet (Ava AG) nightly while sleeping for up to a year or until they became pregnant. In addition to syncing the device to the corresponding smartphone app daily, women also completed an electronic diary about their activities in the past 24 hours. Finally, women took a urinary luteinizing hormone test at several points in a given cycle to determine the close of the fertile window. We assessed phase-based changes in physiological parameters using cross-classified mixed-effects models with random intercepts and random slopes. We then trained a machine learning algorithm to recognize the fertile window.

Results: We have demonstrated that wearable technology can detect significant, concurrent phase-based shifts in WST, heart rate, and respiratory rate (all $P<.001$). HRV and skin perfusion similarly varied across the menstrual cycle (all $P<.05$), although these effects only trended toward significance following a Bonferroni correction to maintain a family-wise alpha level. Our findings were robust to daily, individual, and cycle-level covariates. Furthermore, we developed a machine learning algorithm that can detect the fertile window with 90% accuracy (95% CI 0.89 to 0.92).

Conclusions: Our contributions highlight the impact of artificial intelligence and machine learning's integration into health care. By monitoring numerous physiological parameters simultaneously, wearable technology uniquely improves upon retrospective methods for fertility awareness and enables the first real-time predictive model of ovulation.

(J Med Internet Res 2019;21(4):e13404) doi:10.2196/13404

KEYWORDS

algorithms; fertility/physiology; heart rate; machine learning; menstrual cycle; ovulation detection/methods; respiratory rate; perfusion; skin temperature; wearable electronic devices

Introduction

Background

Wearable sensor technology is evolving rapidly. Primarily providing insights into users' physical activity, these devices have increasingly been adopted in health care settings [1,2]. Sensors embedded in headbands, chest straps, wristwatches, and clothing itself can now track physiological changes that previously required an electroencephalogram, electrocardiograph, electrodermograph, or electromyograph, respectively [2]. Wearable technology renders medical monitoring accessible to everyday consumers. Recent reviews have noted that these devices may allow for greater longitudinal tracking of physiological parameters, enabling users to see personalized patterns developing in the data. To this end, there remains a dearth of research on the applications of wearable technology in health care, especially in women's reproductive health.

Wearable sensor technology that helps women track physiological changes across their menstrual cycle could fill the present gap between high-cost, high-accuracy ovulation detection and free, less-precise fertile window approximation. Transvaginal ultrasound examinations represent the gold standard for ovulation detection; however, they are costly and often not feasible in routine clinical settings [3]. Alternatively, women may consider identifying their fertile window using natural family planning (NFP), based on calendar methods [4-6], basal body temperature (BBT) [6,7], and monitoring the amount and consistency of their cervical mucus as it fluctuates in response to changes in estrogen [8,9]. Less technologically sophisticated and thereby less able to pinpoint ovulation exactly, NFP nevertheless helps women recognize physical symptoms that approximate hormonal, phase-driven changes in their body. Fertility tracking may also involve the use of urine-based luteinizing hormone (LH) kits, which detect the LH surge occurring 24 to 36 hours before ovulation [10,11] and are highly correlated with ovulation detected by ultrasonography [12-14].

Although NFP methods enhance fertility awareness, several reviews have highlighted their shortcomings [6,7,15,16]. Most calendar methods, for example, fail to accommodate natural cycle variation, leading to greater inaccuracy [4,5]. NFP practices relying on physical symptoms similarly suffer from methodological pitfalls. Traditional BBT measurements can be influenced by environmental confounds [17] and cannot prospectively predict the fertile window [4,6,18], whereas cervical mucus monitoring relies on subjective patient interpretation of cervical fluid [7]. Finally, urinary LH tests prospectively identify only the last half of the fertile window [7]; women employing this method to achieve pregnancy risk missing the days with the highest probability of conception, which typically occur before a detectable LH surge [4,7]. Critically, NFP methods require sufficient education for correct application, with their success dependent on user motivation and compliance [19].

Advances in mobile phone technology have hinted at the advantages inherent in clinical applications of wearable technology. Smartphone apps designed to facilitate menstrual cycle tracking have simplified and combined NFP methods. Natural Cycles, for example, relies on the calendar method and BBT together to estimate peak fertility [20,21]. Enabling women to track their cycles from home, smartphone apps nevertheless range in accuracy [22]. At best, they can approximate the fertile window using in-app calculations, thereby removing human interpretation and error; however, recent studies citing the incremental improvements of app-based NFP acknowledge their usefulness would be further heightened through biofeedback [23].

Designed to measure and record physiological parameters, wearable technology seems well-suited to address current limitations in traditional and app-based NFP. Their noninvasive nature allows for the convenient, continuous monitoring of multiple parameters simultaneously, resulting in large datasets and individualized pattern tracking via machine learning [2]. For a fraction of the time, cost, and effort, wearables have the potential to reproduce previous findings, demonstrating the correlation between physiological parameters and the menstrual cycle [17,24]. Spontaneously menstruating women show natural variations in body temperature [25], cardiovascular function [26-28], respiratory rate [29,30], and skin perfusion [31,32], depending on their cycle phase. To document these effects, most previous research has required hospital-grade medical equipment (eg, ultrasound machines [33]). Initial research on a wearable fertility tracker, however, has demonstrated that wrist skin temperature (WST) across the menstrual cycle mirrors BBT-measured phase-based changes. Unlike traditional BBT charting, the correlation was robust to potential confounds [17]. A follow-up study demonstrated that heart rate also serves as a reliable, prospective parameter for cycle tracking [24]. Other devices worn on the wrist, under the arm, or in the ear similarly strive to detect menses and the fertile window through monitoring nightly changes in 1 or 2 physiological changes (eg, core body temperature or heart rate [34-37]); to date, peer-reviewed clinical evidence of their efficacy has not yet been published. Wearables have the potential to improve upon digital calendar methods and BBT by measuring multiple physiological parameters continuously, honing in on an even more precise estimation of ovulation [5,6].

Objectives

Focusing on the application of wearable devices to women's reproductive health, this study represents the first research to track multiple physiological changes concurrently across the menstrual cycle. Using a clinical sample, we aimed to analyze phase-based differences in skin temperature, heart rate, respiratory rate, perfusion, and heart rate variability (HRV). We also probed the robustness of wearable technology, considering the effect of daily, individual, and cycle-level factors on menstrual phase detection. Finally, drawing on insight from artificial intelligence and machine learning, we sought to develop an algorithm that could identify the 6-day fertile window in real time.

Methods

Study Design

Interested in understanding wearable technology's potential as a fertility aid device, we conducted a correlational prospective cohort study enrolling conception-seeking women. Our dependent variables included nightly physiological readings of each parameter, as predicted by the menstrual cycle phase. In addition, we controlled for participant's age, body mass index (BMI; kg/m^2), and other environmental factors that could affect a woman's heart rate, respiratory rate, HRV, WST, or skin perfusion.

Participants

In total, 237 women participated in our study. Previous research has demonstrated the difficulty in assessing the necessary sample size to achieve adequate statistical power in multilevel modeling; namely, power analyses require knowing a priori the effect size of interest, each random effect's variance, covariance estimates for random effects, regression coefficients, and the number of Level 1 groups (eg, how many total cycles and days per cycle each woman will record) [38,39]. As we could not know beforehand the length of each woman's cycle, we sought to recruit a conservative number of participants based on the sample size used in other comparable clinical studies (eg, ranging from 91 women [17] to 317 women [20]).

We recruited participants via flyers hanging in Zurich-area hospitals and private gynecological offices. In addition, we took out a Facebook advertisement targeting Zurich-area women. Both the Web advertisement and paper flyers directed interested individuals to a website where they were asked to complete an entry questionnaire to evaluate inclusion and exclusion criteria, which were established before the initial study enrollment. To meet eligibility criteria, women had to be aged between 18 and 40 years, have regular menstrual cycles (28 [SD 4] days in length), and be trying to conceive. Individuals who reported doing any hormone therapy currently, had health-related issues that affected their menstrual cycle, were on medications or other substances that could interfere with their menstrual cycle or the physiological parameters investigated, traveled frequently across time zones, had a sleeping disorder, and/or had been trying unsuccessfully to become pregnant for more than a year were excluded from the study. Information on each participant's weight and height, used to calculate BMI, was also collected.

Eligible individuals were then contacted by the research team and invited to attend an initial meeting at the Department of Reproductive Medicine at the University Hospital, Zurich. A member of the research team met with each person for at least 30 min to verify eligibility and discuss the study protocol. Individuals had up to a week to provide written informed consent, at which time they were equipped with the necessary study materials; informed consent was obtained from all subjects before their study involvement. Participants reported for in-person follow-up appointments after 3, 6, and 9 subsequent menstrual cycles. Each participant remained enrolled in the study for up to a year or until becoming pregnant. Women who returned study materials at the end of their cycle measurements received 120 Swiss francs as compensation for their participation.

Study Protocol

During their initial meeting with the experimenter, participants received the Web link to a daily diary survey, an electronic wearable to measure physiological parameters while sleeping, and the testing kit for evaluating urinary LH (ClearBlue Advanced Fertility Monitor, SPD Swiss Precision Diagnostics GmbH). Registered with the US Food and Drug Administration as a fertility aid device, the wristworn Ava bracelet (Ava AG) measures 7 physiological parameters simultaneously including WST, heart rate, HRV, respiratory rate, and skin perfusion. The Ava bracelet also measures a user's sleep quantity and sleep quality. These variables were not of interest to the research question presented here and thus are not included in our analyses. Study participants were instructed to wear the bracelet on the dorsal side of their wrist nightly while sleeping. The electronic wearable automatically saves physiological information every 10 seconds throughout the night. During their initial appointment, participants were shown how to sync the device with the complementary app on their smartphone and were instructed to do so each morning upon waking.

In addition to syncing their bracelet daily, participants also completed a Web diary entry about their activities during the last 24 hours. Previous research has indicated that engaging in aerobic exercise or consuming caffeine, alcohol, or food in the 3 hours before bed can affect physiological parameters of interest to our study (eg, body temperature [40-43]). To control for potential covariates, we asked participants to indicate whether they, in the 3 hours preceding sleep, had sexual intercourse, exercised heavily, eaten, drank coffee, or consumed alcohol. In addition, we asked if participants had taken a pregnancy test that day and, if so, whether they were pregnant.

Finally, participants tracked and reported their LH peak each cycle using the ClearBlue Advanced Fertility Monitor. An at-home LH test, it has been widely used in previous research to estimate the day of ovulation (OV) and close of the fertile window [12,14]. The ClearBlue Fertility Advanced Fertility Monitor shows a smiling face when it detects LH levels indicating *peak fertility* (typically 1 day before OV [14]). From 5 days after the onset of menses and through OV, participants measured their LH levels each morning; they reported their result in their daily diary entry.

In keeping with previous research [17,24], we divided the menstrual cycle into the following 5 phases: (1) *menstruation*, starting with the first day of menses and lasting 5 days; (2) the *follicular phase***,** starting on the first day post menses and lasting through 6 days before ovulation (OV –6); (3) the *fertile window***,** starting 5 days before ovulation and lasting through ovulation (OV –5 to OV); (4) the *early luteal phase*, starting 1 day after ovulation through a week after ovulation (OV+1 to OV+7); and

(5) the *late luteal phase*, starting 8 days after ovulation (OV+8) and lasting through the day before the onset of menses.

All research was performed in accordance with the Declaration of Helsinki. The clinical protocol was reviewed and approved by the Cantonal Ethics Committee Zurich, Switzerland (BASEC -Nr 2016-02241). It was registered in ClinicalTrials.gov under the identifier NCT03161873 as well as with the Swiss Federal Complementary Database (Portal) before data collection. Informed consent was obtained from all subjects before their study involvement.

Statistical Analysis

We conducted all data processing and analysis using R (v3.5.1) and Python 3.5. To account for the variation in physiological parameters that arises from sleep onset and awakening [44], we excluded the first 90 and the last 30 min of each night's data a priori. In addition, each parameter underwent locally estimated scatterplot smoothing (LOESS smoothed) before statistical analysis, thereby reducing artificial fluctuations owing to potential measurement error and consistent with best practices [45]. We tested the underlying assumptions of multilevel modeling by examining the residuals of the 5 base models and plotting their respective q-q plots [46].

Next, to probe whether physiological parameters changed across the menstrual cycle, we ran a series of multilevel models with random slopes and random intercepts. Our data were inherently structured, with nightly measurements nested within menstrual cycles and menstrual cycles nested within participants. Thus, we specified cross-classification in our models. The first random effects term specified participant identification number as the random intercept and the cycle number from which the observation was drawn as the random slope. In the second random effects term, we specified the cycle number as the random intercept and the phases of the menstrual cycle as the random slopes. We optimized the model fit via Residual Maximum Likelihood and Satterthwaite degrees of freedom. Specifically, the R packages (R Foundation for Statistical Computing) lme4 [47], lmerTest [48], optimx [49], and multcomp [50] enabled us to test phase-based changes in physiology across the menstrual cycle. When possible, we chose the model using the percentile of data (stable maxima) with the lowest kurtosis and best fit. When missing data rendered it impossible to compare the fit of 2 models, we conservatively chose the model more closely approximating the median observations (eg, the 70th percentile over the 90th percentile).

Given the large number of covariate models we were testing, we implemented a Bonferroni correction to ensure that the family-wise alpha level did not rise above .05. We divided the desired alpha level of .05 by the total number of models we tested (n=50) to arrive at a revised significance level of less than or equal to .001. We adjusted our definition of marginal significance in turn to comprise an alpha value ranging from .05 (the desired family-wise significance level) to .001. We used the Bonferroni-corrected significance level throughout the paper.

Creation of the Fertility Algorithm

After retrospectively analyzing the clinical data, we turned to techniques from machine learning to develop an algorithm for predicting and detecting in real time a woman's fertile window. We used a cycle-based, random 75:25 split for the training and testing datasets with each user belonging to only 1 category; the training dataset consisted of physiological observations from 186 users across 499 cycles, whereas the validation dataset initially contained data from 51 users across 166 cycles. We then trained a random forest with 1000 trees and a max feature parameter of 3 on the training dataset, using the Python module sklearn.ensemble.RandomForestClassifier [51] and the setting max_features=3. We provided 11 input features engineered from the base physiological signals including heart rate, breathing rate, WST, and HRV. We used 3 classes for the initial classification: follicular phase, fertile window, and luteal phase; whereas in our clinical analysis we removed cycles with 20% or more missing data, we kept all cycles in our training dataset. For cycles where participants had synced their data nightly at least 80% of the time, our model used those features in estimating the fertile window. For cycles where nightly data were synced less than 80% of the time, however, the algorithm instead predicted the upcoming fertile window based on the user's previous cycle length and length of their typical luteal phase.

Following the fertility algorithm's training, we tested it using the validation dataset to determine its performance. We removed cycles where women had synced less than 80% of the days from the validation dataset before calculating the reported performance metrics, reflecting the manufacturer's instructions for recommended use. This left us with a validation dataset comprised of 85 cycles spread across 24 users. Interested in our algorithm's ability to correctly predict the fertile window, we then grouped the follicular and luteal phase classifications together into a single nonfertile comparison group. This reclassification allowed us to calculate the algorithm's overall accuracy and *F* score. The *F* score serves as a measure of an algorithm's effectiveness, computed by taking the harmonic average of the mean precision and recall metrics [52].

Results

Physiological Changes Across the Menstrual Cycle

Population Characteristics

From the initially recruited 237 participants, we excluded 44 women's data. In total, 25 participants could not confirm an LH surge in any cycle. In addition, 5 participants had only irregularly long or short menstrual cycles during the course of our study, thereby not meeting the inclusion criteria. Finally, in keeping with previous research and best practices for maximizing fertility prediction algorithms [6,40-43], we also excluded data from 14 women who reported measurements and synchronized their bracelet with the cellphone app less than 80% of days in the cycle.

The final sample included 1194 cycles spread across 193 participants (mean 33.02 [SD 3.68] years); of the 1194 recorded cycles, only 708 met the inclusion criteria for analysis (ie, participants synced their device with the app ≥80% of cycle days and recorded a positive LH test). On average, participants recorded 3.57 analyzable cycles (SD 2.41), with a mean cycle length of 28.21 (SD 2.87) days and a mean BMI of 22.70 kg/m^2 (SD 3.40). Although some women discontinued study participation after a few cycles to pursue *in vitro* fertilization (n=2) or because they no longer wanted to conceive (n=13), no women cited discomfort from the device as a reason for discontinuation.

Physiological Changes in Relation to Cycle Phases

We observed significant changes in physiological parameters as captured by the wearable device across the menstrual cycle. Significant findings from the phase-based analysis of each physiological parameter are reported below, with the fixed effects presented in Table 1; the cross-classified full model for each physiological parameter, including random effects, can be found in Multimedia Appendix 1. In addition, changes in physiological parameters across the menstrual cycle are presented concurrently in Figure 1. Across all models, the menstrual phase served as the reference group, with each of the other 4 phases compared with it directly. Furthermore, we tested each base model separately before including potential covariates; unless otherwise noted, individual (eg, BMI and age), cycle-specific (eg, duration), and/or daily (eg, drinking alcohol in the 3 hours preceding sleep) covariates did not change the direction or significance of phase-based effects.

Table 1. Multilevel linear mixed models reveal the relationship between menstrual phase and physiological parameters.

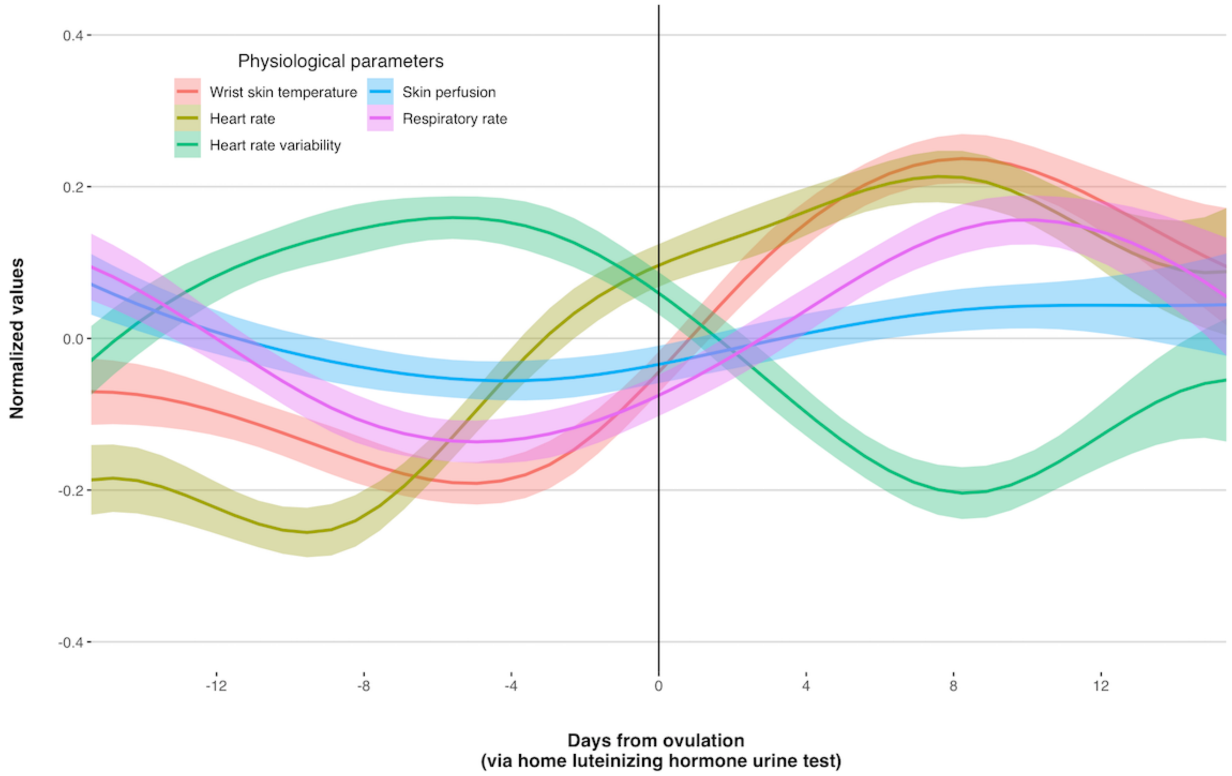
Predictors	Wrist skin temperature[a]	Heart rate[a]	Heart rate variability[a]	Respiratory rate[a]	Skin perfusion[a]
Intercept	34.08[b] (0.08)	58.62[b] (0.44)	1.70[b] (0.10)	16.92[b] (0.14)	1306.54[b] (75.39)
Cycle phase					
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	−0.24[b] (0.02)	−1.54[b] (0.20)	0.11[c] (0.03)	−0.39[b] (0.04)	−44.32[c] (14.80)
Fertile	−0.25[b] (0.03)	−0.03 (0.26)	0.08[c] (0.03)	−0.48[b] (0.04)	−73.58[c] (18.25)
Early luteal	0.01 (0.02)	2.01[b] (0.20)	−0.11 (0.05)	−0.20[b] (0.04)	−12.02 (18.49)
Late luteal	0.20[b] (0.02)	2.46[b] (0.29)	−0.20[c] (0.04)	0.22[b] (0.03)	51.33 (26.96)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] *P*<.001 with a Bonferroni correction.

[c] *P*<.05 with a Bonferroni correction.

Figure 1. Wearable technology can detect changes in 5 physiological parameters across the menstrual cycle. The smoothed, normalized values of each physiological parameter (with 95% CIs) collapsed across individuals (n=193) and cycles (n=708) are shown, centered around participant-reported luteinizing hormone peak.



Wrist Skin Temperature

Regressing nightly WST on menstrual phases using the 50th percentile data revealed women had a significantly lower WST in the follicular phase (mean 33.87 °C [SD 0.84]; $t_{5.06}=-11.12$; $P<.001$) and fertile window (mean 33.88 °C [SD 0.78]; $t_{4.96}=-8.97$; $P<.001$) compared with menses (mean 34.11 °C [SD 0.84]). WST was also significantly higher in the late luteal (mean 34.32 °C [SD 0.82];$t_{14.80}=10.96$; $P<.001$) phase compared with menses.

Controlling for individual, cycle-level, or daily covariates did not change the effect of menstrual phase on WST. In general, women with higher BMIs had significantly lower nightly WST ($t_{184.69}=-3.70$; $P<.001$). Compared with nights when she did not eat or had only a small meal in the 3 hours before sleep, a woman had significantly lower WST after eating a medium- ($t_{16220}=-3.58$; $P<.001$) or large-sized meal ($t_{16240}=-5.10$; $P<.001$; see Table 2).

Heart Rate

There was a significant effect of cycle phase on average nightly heart rate in the data drawn from the 30th percentile. In particular, heart rate was significantly lower in the follicular phase (mean 56.56 [SD 6.29] beats per minute [bpm]) compared with the menstrual phase (mean 57.92 [SD 6.27] bpm; $t_{8.43}=-7.66$; $P<.001$). Nightly heart rate was significantly higher in the early (mean 59.98 [SD 6.50] bpm; $t_{5.74}=9.93$; $P<.001$) and late luteal phases (mean 60.47 [SD 6.45] bpm; $t_{6.02}=8.53$; $P<.001$) than in menses, however.

Daily and cycle-level variables significantly affected nightly heart rate, over and above the effect of the menstrual phase. Women who ate a large-sized meal ($t_{16220}=8.16$; $P<.001$), drank at least 1 serving of alcohol (1 to 4 units, $t_{16250}=12.43$; $P<.001$; ≥5 units, $t_{16250}=18.28$; $P<.001$), and/or exercised for at least 60 min ($t_{16280}=4.84$; $P<.001$) in the 3 hours before sleep had significantly higher heart rates on a given night. Finally, during a longer cycle, women were significantly more likely to have an increased heart rate on a given night ($t_{14000}=5.44$; $P<.001$). Inclusion of these covariates in the model did not affect the direction or significance of phase-based effects (see Table 3).

Table 2. Multilevel linear mixed models reveal the relationship between menstrual phase, covariates, and wrist skin temperature.

Predictors	Model 1[a]	Model 2[a]	Model 3[a]	Model 4[a]	Model 5[a]	Model 6[a]	Model 7[a]	Model 8[a]	Model 9[a]
Intercept	34.12[b] (0.08)	35.40[b] (0.36)	34.09[b] (0.08)	34.09[b] (0.08)	34.09[b] (0.08)	34.09[b] (0.08)	33.32[b] (0.48)	34.24[b] (0.10)	35.01[b] (0.64)
Cycle phase									
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	–0.24[b] (0.03)	–0.23[b] (0.03)	–0.24[b] (0.03)	–0.24[b] (0.03)	–0.24[b] (0.02)	–0.24[b] (0.03)	–0.24[b] (0.03)	–0.24[b] (0.03)	–0.23[b] (0.03)
Fertile	–0.26[b] (0.03)	–0.25[b] (0.03)	–0.26[b] (0.03)	–0.26[b] (0.03)	–0.25[b] (0.03)	–0.26[b] (0.03)	–0.26[b] (0.03)	–0.26[b] (0.03)	–0.24[b] (0.03)
Early luteal	0.00 (0.02)	0.01 (0.03)	0.00 (0.02)	0.00 (0.02)	0.01 (0.02)	0.00 (0.02)	0.00 (0.03)	0.00 (0.02)	0.02 (0.03)
Late luteal	0.19[b] (0.02)	0.20[b] (0.02)	0.19[b] (0.02)	0.19[b] (0.02)	0.19[b] (0.02)	0.19[b] (0.02)	0.20[b] (0.02)	0.19[b] (0.02)	0.20[b] (0.02)
Meal[c]									
Small or no food	Reference group	—[d]	—	—	—	—	—	—	Reference group
Medium meal	–0.03[b] (0.01)	—	—	—	—	—	—	—	–0.03[b] (0.01)
Large meal	–0.05[b] (0.01)	—	—	—	—	—	—	—	–0.05[b] (0.01)
Body mass index (kg/m^2)	—	–0.06[b] (0.02)	—	—	—	—	—	—	–0.05[b] (0.02)
Coffee[c]	—	—	0.00 (0.01)	—	—	—	—	—	0.01 (0.01)
Exercise[c]									
No exercise	—	—	—	Reference group	—	—	—	—	Reference group
<60 min	—	—	—	0.01 (0.01)	—	—	—	—	0.01 (0.01)
>60 min	—	—	—	0.02 (0.01)	—	—	—	—	0.01 (0.01)
Sexual intercourse[c]	—	—	—	—	–0.02[e] (0.01)	—	—	—	–0.02[e] (0.01)
Alcohol[c]									
No alcohol	—	—	—	—	—	Reference group	—	—	Reference group
1-4 units	—	—	—	—	—	–0.01 (0.01)	—	—	0.00 (0.01)
≥5 units	—	—	—	—	—	0.05[e] (0.02)	—	—	0.07[d] (0.02)
Age (years)	—	—	—	—	—	—	0.02 (0.01)	—	0.02 (0.01)
Cycle duration	—	—	—	—	—	—	—	–0.01[e] (0.00)	–0.01[e] (0.00)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] *P*<.001 with a Bonferroni correction.

[c] Within the 3 hours preceding the onset of sleep.

[d] Indicates a predictor was not considered in a given model.

[e] *P*<.05 with a Bonferroni correction.

Table 3. Multilevel linear mixed models reveal the relationship between menstrual phase, covariates, and heart rate.

Predictors	Model 1[a]	Model 2[a]	Model 3[a,b]	Model 4[a,b]	Model 5[a]	Model 6[a]	Model 7[a]	Model 8[a,b]	Model 9[a]
Intercept	58.40[c] (0.44)	54.02[c] (2.86)	58.58[c] (0.44)	58.53[c] (0.44)	58.60[c] (0.44)	58.38[c] (0.44)	60.37[c] (3.81)	56.51[c] (0.58)	53.52[c] (5.06)
Cycle phase									
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	–1.57[c] (0.21)	–1.57[c] (0.21)	–1.29[c] (0.07)	–1.30[c] (0.07)	1.57[c] (0.21)	–1.58[c] (0.22)	–1.55[c] (0.20)	–1.31[c] (0.07)	–1.62[c] (0.22)
Fertile	–0.04 (0.25)	–0.03 (0.23)	0.19[d] (0.07)	0.19[d] (0.07)	–0.04 (0.25)	–0.04 (0.24)	–0.02 (0.24)	0.19[d] (0.06)	–0.02 (0.24)
Early luteal	1.95[c] (0.19)	1.98[c] (0.16)	2.01[c] (0.06)	2.01[c] (0.06)	1.95[c] (0.18)	1.98[c] (0.18)	1.96[c] (0.17)	2.01[c] (0.06)	2.00[c] (0.17)
Late luteal	2.39[c] (0.24)	2.40[c] (0.24)	2.37[c] (0.07)	2.37[c] (0.07)	2.40[c] (0.24)	2.41[c] (0.25)	2.40[c] (0.24)	2.37[c] (0.07)	2.40[c] (0.25)
Meal[d]									
Small or no food	Reference group	—[e]	—	—	—	—	—	—	Reference group
Medium meal	0.08 (0.06)	—	—	—	—	—	—	—	0.05 (0.06)
Large meal	0.56[c] (0.07)	—	—	—	—	—	—	—	0.31[c] (0.07)
Body mass index (kg/m^2)	—	0.21 (0.12)	—	—	—	—	—	—	0.20 (0.13)
Coffee[d]	—	—	0.02 (0.08)	—	—	—	—	—	–0.14 (0.08)
Exercise[d]									
No exercise	—	—	—	Reference group	—	—	—	—	Reference group
<60 min	—	—	—	0.03 (0.06)	—	—	—	—	0.05 (0.06)
>60 min	—	—	—	0.33[c] (0.07)	—	—	—	—	0.36[c] (0.07)
Sexual intercourse[d]	—	—	—	—	0.03 (0.06)	—	—	—	–0.04 (0.06)
Alcohol[d]									
No alcohol	—	—	—	—	—	Reference group	—	—	Reference group
1-4 units	—	—	—	—	—	0.62[c] (0.05)	—	—	0.53[c] (0.05)
≥5 units	—	—	—	—	—	2.82[c] (0.15)	—	—	2.70[c] (0.16)
Age (years)	—	—	—	—	—	—	–0.05 (0.11)	—	–0.05 (0.12)
Cycle duration	—	—	—	—	—	—		0.07[c] (0.01)	0.06[c] (0.01)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] The model would not converge with the cross-classification term, so only random intercepts were included.

[c] *P*<.001 with a Bonferroni correction.

[d] *P*<.05 with a Bonferroni correction.

[d] Within the 3 hours preceding the onset of sleep.

[e] Indicates a predictor was not considered in a given model.

Heart Rate Variability

Analysis of data from the 90th percentile revealed a marginally significant effect of cycle phase on the criterion. The average nightly HRV ratio was higher in the follicular phase (mean 1.86 [SD 0.91]; $t_{4.10}$=3.28; P=.03) and fertile window (mean 1.86 [SD 0.91]; $t_{3.49}$=3.37; P=.03) than in the menstrual phase (mean 1.78 [SD 0.88]). The HRV ratio dipped during the luteal phase, tending to be lower on a given night in the late luteal phase (mean 1.62 [SD 0.79]) than during menses ($t_{4.38}$=–5.50; P=.004). Not meeting the more conservative Bonferroni corrected alpha level of .001, however, these phase-based differences only trended toward significance.

A covariate analysis revealed only a significant effect of daily-level predictors on HRV, over and above the effects of the menstrual phase. In particular, women had lower HRV ratios on a given night if they had eaten a large-sized meal (t_{16280}=–4.35; P<.001) compared with nights where they fasted or ate only a small meal in the 3 hours before bed. Nevertheless, the phase-based trends in HRV ratio remained robust; compared with menses, the HRV ratio was higher in the follicular phase and fertile window, but lower during the luteal phase (see Table 4).

Respiratory Rate

Examining data from the 90th percentile of nightly observations, respiratory rate was significantly lower in the follicular phase (mean 16.57 [SD 2.06] breaths/min; $t_{8.55}$=–9.96; P<.001), fertile window (mean 16.40 [SD 2.00] breaths/min; $t_{9.52}$=–11.59; P<.001), and early luteal phase (mean 16.68 [SD 1.96] breaths/min; $t_{9.39}$=–5.44; P<.001) compared with menses (mean 16.86 [SD 2.03] breaths/min). Finally, during the late luteal phase, a woman's respiratory rate was significantly faster (mean 17.04 [SD 1.97] breaths/min; $t_{7.52}$=6.74; P<.001) than during menses.

Eating a large meal (t_{16210}=5.39; P<.001) or consuming alcohol (1 to 4 units, t_{16200}=8.86; P<.001; ≥5 units, t_{16210}=12.56; P<.001) in the 3 hours preceding sleep was associated with a significant increase in nightly respiratory rate, over and above the effects of the menstrual phase. When considered alongside all measured covariates in a single model, having sexual intercourse significantly decreased a woman's nightly respiratory rate compared with nights where she was abstinent, over and above the effects of the menstrual phase (t_{15250}=–3.50; P<.001). Accounting for the effects of individual covariates, however, did not alter the direction or significance of the menstrual phase on respiratory rate (see Table 5).

Skin Perfusion

Mixed-effects modeling with random slopes and random intercepts using the 90th percentile data revealed a marginally significant effect of cycle phase on skin perfusion; on average, wrist skin was less perfuse during the follicular phase (mean 1438.78 [SD 599.43]; $t_{6.33}$=–3.00; P=.02) and the fertile window (mean 1384.02 [SD 577.19]; $t_{7.37}$=–4.03; P=.004) compared with menses (mean 1431.53 [SD 592.22]).

Inclusion of covariates in the base model did not affect the direction of the relationship between menses and the follicular phase or menses and the fertile window (see Table 6). Drinking 1 to 4 units of alcohol ($t_{16247.87}$=–4.11; P<.001) in the 3 hours before sleep significantly decreased nightly skin perfusion, over and above the marginally significant effects of cycle phase. In addition, eating a medium- ($t_{16237.84}$=–3.67; P<.001) or large-sized meal ($t_{16254.02}$=–5.85; P<.001) in the 3 hours preceding sleep significantly reduced nightly perfusion, compared with when a woman fasted or had only a small meal before bed.

Table 4. Multilevel linear mixed models reveal the relationship between the menstrual phase, covariates, and heart rate variability.

Predictors	Model 1[a]	Model 2[a]	Model 3[a]	Model 4[a,b]	Model 5[a,b]	Model 6[a,b]	Model 7[a]	Model 8[a]	Model 9[a,b]
Intercept	1.75[c] (0.10)	1.51[c] (0.37)	1.73[c] (0.10)	1.74[c] (0.09)	1.74[c] (0.09)	1.75[c] (0.09)	0.77 (0.49)	1.70[c] (0.11)	0.32 (0.64)
Cycle phase									
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	0.12[d] (0.03)	0.12[d] (0.03)	0.12[d] (0.03)	0.11[c] (0.01)	0.11[c] (0.01)	0.11[c] (0.01)	0.12[d] (0.03)	0.17[d] (0.03)	0.11[c] (0.01)
Fertile	0.08[d] (0.02)	0.07[d] (0.02)	0.08[d] (0.02)	0.07[c] (0.01)	0.07[c] (0.01)	0.07[c] (0.01)	0.07[d] (0.02)	0.08[d] (0.02)	0.06[c] (0.01)
Early luteal	–0.12[d] (0.04)	–0.11[d] (0.04)	–0.11[d] (0.04)	–0.11[c] (0.01)	–0.11[c] (0.01)	–0.11[c] (0.01)	–0.12[d] (0.04)	–0.12[d] (0.04)	–0.11[c] (0.01)
Late luteal	–0.20[d] (0.04)	–0.20[d] (0.04)	–0.12[d] (0.01)	–0.19[c] (0.01)	–0.19[c] (0.01)	–0.19[c] (0.01)	–0.20[d] (0.04)	–0.20[d] (0.04)	–0.19[c] (0.01)
Meal[e]									
Small or no food	Reference group	—[f]	—	—	—	—	—	—	Reference group
Medium meal	–0.02[d] (0.01)	—	—	—	—	—	—	—	–0.02[e] (0.01)
Large meal	–0.04[c] (0.01)	—	—	—	—	—	—	—	–0.03[e] (0.01)
Body mass index (kg/m^2)	—	0.01 (0.02)	—	—	—	—	—	—	0.01 (0.02)
Coffee[e]	—	—	–0.03[d] (0.01)	—	—	—	—	—	–0.02 (0.01)
Exercise[e]									
No exercise	—	—	—	Reference group	—	—	—	—	Reference group
<60 min	—	—	—	0.01 (0.01)	—	—	—	—	0.01 (0.01)
>60 min	—	—	—	0.01 (0.01)	—	—	—	—	0.00 (0.01)
Sexual intercourse[e]	—	—	—	—	–0.02 (0.01)	—	—	—	–0.01 (0.01)
Alcohol[e]									
No alcohol	—	—	—	—	—	Reference group	—	—	Reference group
1-4 units	—	—	—	—	—	–0.02[d] (0.01)	—	—	–0.01(0.01)
≥5 units	—	—	—	—	—	–0.06[d] (0.02)	—	—	–0.05(0.02)
Age (years)	—	—	—	—	—	—	0.03[d] (0.01)	—	0.03[d] (0.01)
Cycle duration	—	—	—	—	—	—	—	0.00 (0.00)	0.00 (0.00)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] The model would not converge with the cross-classification term, so only random intercepts were included.

[c] *P*<.001 with a Bonferroni correction.

[d] *P*<.05 with a Bonferroni correction.

[e] Within the 3 hours preceding the onset of sleep.

[f] Indicates a predictor was not considered in a given model.

Table 5. Multilevel linear mixed models reveal the relationship between menstrual phase, covariates, and respiratory rate.

Predictors	Model 1[a]	Model 2[a]	Model 3[a]	Model 4[a]	Model 5[a]	Model 6[a]	Model 7[a]	Model 8[a]	Model 9[a]
Intercept	16.89[b] (0.14)	15.39[b] (0.97)	16.91[b] (0.14)	16.92[b] (0.14)	16.92[b] (0.14)	16.88[b] (0.14)	17.82[b] (1.28)	16.88[b] (0.17)	16.15[b] (1.72)
Cycle phase									
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	–0.38[b] (0.04)	–0.39[b] (0.05)	–0.38[b] (0.04)	–0.38[b] (0.04)	–0.38[b] (0.04)	–0.38[b] (0.05)	–0.38[b] (0.05)	–0.38[b] (0.04)	–0.39[b] (0.05)
Fertile	–0.48[b] (0.04)	–0.48[b] (0.04)	–0.47[b] (0.04)	–0.48[b] (0.04)	–0.46[b] (0.04)	–0.48[b] (0.04)	–0.47[b] (0.04)	–0.47[b] (0.04)	–0.46[b] (0.04)
Early luteal	–0.21[b] (0.04)	–0.22[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)	–0.21[b] (0.04)
Late luteal	0.21[b] (0.03)	0.21[b] (0.03)	0.22[b] (0.03)	0.22[b] (0.03)	0.22[b] (0.03)	0.21[b] (0.03)	0.22[b] (0.03)	0.22[b] (0.03)	0.22[b] (0.03)
Meal[c]									
Small or no food	Reference group	—[d]	—	—	—	—	—	—	Reference group
Medium meal	0.00 (0.02)	—	—	—	—	—	—	—	–0.01 (0.02)
Large meal	0.09[b] (0.02)	—	—	—	—	—	—	—	0.04[e] (0.02)
Body mass index (kg/m^2)	—	0.07 (0.04)	—	—	—	—	—	—	0.06 (0.04)
Coffee[c]	—	—	.06[e] (0.02)	—	—	—	—	—	0.03 (0.02)
Exercise[c]									
No exercise	—	—	—	Reference group	—	—	—	—	Reference group
<60 min	—	—	—	–0.01 (0.02)	—	—	—	—	–0.01 (0.02)
>60 min	—	—	—	0.03 (0.02)	—	—	—	—	0.03 (0.02)
Sexual intercourse[c]	—	—	—	—	–0.04[e] (0.01)	—	—	—	–0.05[b] (0.02)
Alcohol[c]									
No alcohol	—	—	—	—	—	Reference group	—	—	Reference group
1-4 units	—	—	—	—	—	0.11[b] (0.01)	—	—	0.10[b] (0.01)
>5 units	—	—	—	—	—	0.48[b] (0.04)	—	—	0.46[b] (0.04)
Age (years)	—	—	—	—	—	—	–0.03 (0.04)	—	–0.02 (0.04)
Cycle duration	—	—	—	—	—	—	—	0.00 (0.00)	0.00 (0.00)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] *P*<.001 with a Bonferroni correction.

[c] Within the 3 hours preceding the onset of sleep.

[d] Indicates a predictor was not considered in a given model.

[e] *P*<.05 with a Bonferroni correction.

Table 6. Multilevel linear mixed models reveal the relationship between menstrual phase, covariates, and skin perfusion.

Predictors	Model 1[a]	Model 2[a,b]	Model 3[a]	Model 4[a,b]	Model 5[a]	Model 6[a]	Model 7[a,b]	Model 8[a]	Model 9[a,b]
Intercept	1338.17[c] (72.65)	811.67[d] (244.86)	1308.46[c] (72.58)	1316.50[c] (74.72)	1308.46[c] (72.59)	1315.27[c] (72.86)	1629.12[c] (319.49)	1325.16[c] (85.35)	1080.00[d] (426.30)
Cycle phase									
Menstrual	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group	Reference group
Follicular	–49.62[d] (15.07)	–27.30[c] (8.28)	–49.94[d] (15.20)	–30.93[c] (8.27)	–48.22[d] (15.12)	–50.23[d] (15.28)	–30.89[c] (8.32)	–49.29[d] (15.09)	–26.76[d] (8.35)
Fertile	–68.23[d] (16.94)	–45.20[c] (7.69)	–68.77[d] (17.03)	–48.69[c] (7.70)	–65.00[d] (16.99)	–68.17[d] (16.74)	–48.91[c] (7.74)	–68.04[d] (16.87)	–43.89[c] (7.96)
Early luteal	–6.37 (17.48)	–15.96[d] (7.50)	–7.14 (17.37)	–12.76 (7.51)	–5.64 (17.54)	–6.88 (17.56)	–14.68 (7.54)	–7.06 (17.68)	–1.53[d] (7.53)
Late luteal	60.82 (28.62)	34.01[c] (8.09)	60.73 (28.60)	33.85[c] (8.11)	61.94 (28.71)	60.18 (28.75)	33.83[c] (8.14)	61.94 (29.06)	33.43[c] (8.13)
Meal[e]									
Small or no food	Reference group	—[f]	—	—	—	—	—	—	Reference group
Medium meal	–26.43[c] (7.21)	—	—	—	—	—	—	—	–20.55[d] (7.29)
Large meal	–47.36[c] (8.09)	—	—	—	—	—	—	—	–40.31[c] (8.52)
Body mass index (kg/m^2)	—	22.08[d] (10.21)	—	—	—	—	—	—	21.19[d] (10.33)
Coffee[e]	—	—	–7.00 (9.77)	—	—	—	—	—	–0.04 (9.77)
Exercise[e]									
No exercise	—	—	—	Reference group	—	—	—	—	Reference group
<60 min	—	—	—	–17.81[d] (7.46)	—	—	—	—	–18.56[d] (7.45)
>60 min	—	—	—	11.12 (8.03)	—	—	—	—	7.65 (8.11)
Sexual intercourse[e]	—	—	—	—	–11.42 (7.00)	—	—	—	–4.90 (7.03)
Alcohol[e]									
No alcohol	—	—	—	—	—	Reference group	—	—	Reference group
1-4 units	—	—	—	—	—	–24.71[c] (6.01)	—	—	–16.89[d] (6.33)
>5 units	—	—	—	—	—	–17.64 (18.50)	—	—	–13.31 (18.56)
Age (years)	—	—	—	—	—	—	–9.56 (9.36)	—	–6.59 (9.49)
Cycle duration	—	—	—	—	—	—	—	–0.62 (1.59)	–0.01 (1.61)

[a] Unstandardized b-coefficient values reported, with SEs in parentheses.

[b] The model would not converge with the cross-classification term, so only random intercepts were included.

[c] *P*<.001 with a Bonferroni correction.

[d] *P*<.05 with a Bonferroni correction.

[e] Within the 3 hours preceding the onset of sleep.

[f] Indicates a predictor was not considered in a given model.

Fertility Prediction Algorithm Performance

Having demonstrated the significant changes in physiological parameters across the menstrual cycle, we proceeded to develop a predictive real-time model for detection of the fertile window. Employing an ensemble tree-based machine learning method resulted in good separation between the different phases of the menstrual cycle, with some phases easier to isolate than others (see Multimedia Appendix 2 for the confusion matrix). The overall method performance based on the *F* score was 0.78 (95% CI 0.74 to 0.82; specificity=0.93, 95% CI 0.92 to 0.94; sensitivity=0.81, 95% CI 0.77 to 0.85); furthermore, the algorithm accurately detected the 6-day fertile window in 90% of cycles (95% CI 0.89 to 0.92; see Multimedia Appendix 3 for full performance metrics). A 2018 review of smartphone apps for tracking the menstrual cycle found that fewer than a third of the 73 apps surveyed (17.3%) could predict a 6-day or smaller fertile window, achieving an accuracy between 11% and 81% (mean 53% [SD 21%]) [53]. In addition to having a shorter fertile window than most of the apps studied, our algorithm had a higher accuracy metric. Although NFP techniques have demonstrated an achieved accuracy of up to 98%, they require assumptions about the next cycle's duration to determine accuracy and do so at the expense of providing a broader fertile window (ranging from 8 to 64 days for the rhythm method) [53]. By drawing on machine learning and users' detailed physiological profiles across multiple cycles, our fertility prediction algorithm achieves higher accuracy than previous smartphone apps and pinpoints a fertile window more precisely than traditional NFP.

Discussion

Principal Findings

Probing wearable technology's ability to monitor multiple physiological parameters concurrently, our study demonstrated how WST, heart rate, HRV, respiratory rate, and skin perfusion vary across the menstrual cycle. In line with previous research [17], we captured a biphasic shift in WST; compared with menses, women had significantly lower WST in the follicular and fertile phase and significantly higher WST in the late luteal phase. The phase-based changes in WST mirror findings from studies using more traditional BBT methods for temperature tracking [4,6,31]. Wearable technology renders similar readings and conclusions as BBT in a less invasive manner, solving many of BBT's inherent disadvantages (eg, the need to take one's temperature at the same time daily [7]). Our study also upholds conclusions from previous research, which revealed women have significantly higher resting pulse rates [26,28-30,54] and respiratory rates [29,30] during the luteal phase compared with earlier in their menstrual cycle. Although we have previously captured phase-based changes for a single parameter in an ambulatory setting [17,24], we have demonstrated here for the first time that wearable technology can track multiple physiological parameters across the menstrual cycle simultaneously. In turn, this increase in recorded features allowed us to harness machine learning to predict the opening and closing of the fertile window with high accuracy.

Limitations

Not all phase-based changes in physiological parameters manifested as we predicted, however. Previous research considering the effect of the menstrual cycle on perfusion led us to expect significantly less skin perfusion in the luteal and menstrual phases compared with the follicular phase or fertile window [55]. However, our results trended in the opposite direction, with skin perfusion lower during the follicular phase and fertile window compared with menses. Methodological differences in protocol may partially explain the discrepancy between our findings and past work. Although most researchers measured skin perfusion via sensors on a participant's finger or their forearm [55], participants in this study wore our device on the dorsal side of their wrist. Previous research has found that, even within the same study, population, and time frame, conclusions about the physiological changes in skin perfusion across the menstrual cycle may differ depending on where on a participant's finger the measuring instrument was placed [56]. Future research should consider how the location of sensors monitoring peripheral blood flow moderates phase-based differences in skin perfusion.

Like skin perfusion, HRV also showed marginally significant phase-based changes across the menstrual cycle in the opposite direction than anticipated. Multiple studies have reported women have higher HRV ratios during the luteal phase than earlier in their cycle [27,28,54]. In contrast, we found that the HRV ratio increased during the follicular phase compared with menses before decreasing during the luteal phase. Although a definitive explanation would require follow-up studies, we believe this difference may be due to variability in the experimental context. Owing to wearable technology's ambulatory nature, we could measure HRV every 10 seconds throughout the night as participants slept. Previous research, however, required subjects to report to a laboratory or hospital during the day, where experimenters collected HRV measurements for a comparatively brief period (eg, 30 min [29]). In addition to a smaller sampling distribution owing to temporal constraints, the participant's waking state may have contributed to differences in findings. Time awake has been shown to moderate the effect of the menstrual cycle on HRV ratio, with sleep deprivation significantly increasing sympathetic activity during the midfollicular phase [57]. To better understand why our findings trended in the opposite direction than expected, follow-up studies may wish to directly compare phase-based changes in the HRV ratio across participants' sleep and waking states.

An additional limitation of our study was that we computed the fertility prediction algorithm's accuracy based on compliant users who synced their bracelet with the smartphone app at least 80% of the days in a given cycle. Previous research suggests that real-world adherence to reproductive health protocol varies greatly, however. Studies looking at technology-based contraception methods, for example, calculate both the *perfect use* as well as the *typical use* rate of unintended pregnancies, assuming an average user may not follow directions as written all the time (eg, [15,21]). Although we required users to sync their wearable device at least 80% of the days in a given cycle, we attempted to simulate how our algorithm would perform for less compliant users. To do so, we wrote a Python script to randomly remove 10%, 30%, or 50% of the nightly observations from the validation dataset and then recalculated the algorithm's accuracy and key performance metrics for each simulated amount of noncompliance. As may be expected, the fertility prediction algorithm became increasingly less accurate as more observations were deleted from the dataset. Nevertheless, even with 50% of the observations removed, our fertility algorithm accurately predicted the fertile window in more than 86% of cycles (95% CI 0.85 to 0.89) and had an F1 score equal to 0.72 (95% CI 0.68 to 0.76; see Multimedia Appendix 3 for the performance metrics across all degrees of simulated noncompliance). Planned future analyses will consider the effect of protocol adherence on algorithm prediction accuracy among real-world users. Given our simulated results, we expect a relatively high performance even among users who sync less than 80% of the time.

Conclusions

Menstrual cycle tracking has numerous applications in health care, further augmented by the development of wearables. First, identification of the fertile window can aid couples planning a pregnancy. A retrospective survey of women who conceived with the help of a fertility monitor found mistiming sexual intercourse to be a leading reason for unexplained infertility; knowledge about their fertile window allowed 49.5% of women in the sample to conceive the subsequent menstrual cycle [58]. Wearables improve upon this possibility, triangulating the fertile window through continuous, high-frequency measurement of multiple parameters. Detection of the fertile window can also assist women who wish to avoid pregnancy but cannot or do not want to use hormonal contraception [15].

Wearable devices provide women with an accurate, convenient alternative to other methods for predicting the fertile window. We have described the measurement, analysis, and interpretation of menstrual cycle–related physiological changes using a wearable device. In our study, wristworn wearables show similar sensitivity and specificity as more invasive, time intensive NFP methods, including cervical mucus monitoring and LH testing [59]. Wearable technology can provide women with real-time insight into their bodies and menstrual cycles, serving as an at-home educational tool. Furthermore, access to cyclic data via a mobile app enables long-term cycle tracking and can lead to more informed lifestyle and medical decisions. Future research should consider how wearable technology can help elucidate physiological patterns underlying women's health care concerns.

We reproduced the results of previous studies using a more accurate, distributable technology. Significant phase-based differences emerged for 3 of the physiological parameters of interest; the remaining 2 physiological parameters showed similar trends toward phase-based differences, despite a conservative correction to reduce Type I error rates. On the basis of signals collected by the wearable device, we created an algorithm that predicts each woman's most fertile days in a given cycle. Our study suggests wearables' imperviousness to confounding factors; the wristworn device detected changes in the menstrual phase over and above daily, cyclical, or individual level fluctuations in parameters. Wearable sensor technology enables the first real-time predictive model of ovulation and represents a valuable addition to women's health care.

Acknowledgments

The authors would like to thank Florian Kübler and Natascha Hagger for their help in recruiting and equipping study participants. In addition, the authors are grateful to Andres Varags Lugo Cantu for his initial contributions to the fertility algorithm. Finally, the authors would be remiss if they did not thank the Swiss Commission for Technology and Innovation for their grant funding, which helped make this research possible.

Authors' Contributions

BMG ran the statistical analyses on the clinical sample and cowrote the first draft of this paper. MS oversaw the data collection process, ran the initial statistical analysis, and cowrote the first draft of this paper. FD, LF, and MS developed and tested the fertility algorithm. GH contributed to study organization and participant recruitment. BL served as the supervisor and primary investigator for the study. All authors participated in finalization of this paper.

Conflicts of Interest

The authors declare the existence of a financial conflict of interest. BMG, MS, LF, FD, and GH are current or previous employees at Ava AG. BL serves on the Ava AG medical advisory board.

Multimedia Appendix 1

R code output for cross-classified multi-level models of physiological parameters.

[DOCX File, 27KB - jmir_v21i4e13404_app1.docx]

Multimedia Appendix 2

Confusion matrix for the Fertility Prediction Algorithm.

[DOCX File, 18KB - jmir_v21i4e13404_app2.docx]

Multimedia Appendix 3

Real and simulated Fertility Prediction Algorithm performance for compliant and noncompliant participants.

[DOCX File, 28KB - jmir_v21i4e13404_app3.docx]

References

1. Henriksen A, Haugen MM, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
2. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
3. Ecochard R, Boehringer H, Rabilloud M, Marret H. Chronological aspects of ultrasonic, hormonal, and other indirect indices of ovulation. BJOG 2001 Aug;108(8):822-829 [FREE Full text] [doi: 10.1111/j.1471-0528.2001.00194.x] [Medline: 11510707]
4. Stanford JB, White GL, Hatasaka H. Timing intercourse to achieve pregnancy: current evidence. Obstet Gynecol 2002 Dec;100(6):1333-1341. [Medline: 12468181]
5. Fehring RJ. New low- and high-tech calendar methods of family planning. J Midwifery Womens Health 2005;50(1):31-38. [doi: 10.1016/j.jmwh.2004.07.001] [Medline: 15637512]
6. Pallone SR, Bergus GR. Fertility awareness-based methods: another option for family planning. J Am Board Fam Med 2009;22(2):147-157 [FREE Full text] [doi: 10.3122/jabfm.2009.02.080038] [Medline: 19264938]
7. Brezina PR, Haberl E, Wallach E. At home testing: optimizing management for the infertility physician. Fertil Steril 2011 May;95(6):1867-1878. [doi: 10.1016/j.fertnstert.2011.01.001] [Medline: 21324446]
8. Howard MP, Stanford JB. Pregnancy probabilities during use of the Creighton Model Fertility Care System. Arch Fam Med 1999;8(5):391-402. [Medline: 10500511]
9. Arévalo M, Jennings V, Nikula M, Sinai I. Efficacy of the new TwoDay Method of family planning. Fertil Steril 2004 Oct;82(4):885-892. [doi: 10.1016/j.fertnstert.2004.03.040] [Medline: 15482764]
10. Singh M, Saxena BB, Rathnam P. Clinical validation of enzymeimmunoassay of human luteinizing hormone (hLH) in the detection of the preovulatory luteinizing hormone (LH) surge in urine. Fertil Steril 1984 Feb;41(2):210-217. [doi: 10.1016/S0015-0282(16)47592-8] [Medline: 6365598]
11. World Health Organization. Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17 beta, luteinizing hormone, follicle-stimulating hormone, and progesterone. I. Probit analysis. World Health Organization, Task Force on Methods for the Determination of the Fertile Period, Special Programme of Research, Development and Research Training in Human Reproduction. Am J Obstet Gynecol 1980 Oct 15;138(4):383-390. [Medline: 6775535]
12. Guida M, Tommaselli GA, Palomba S, Pellicano M, Moccia G, Di Carlo C, et al. Efficacy of methods for determining ovulation in a natural family planning program. Fertil Steril 1999 Nov;72(5):900-904. [doi: 10.1016/S0015-0282(99)00365-9] [Medline: 10560997]
13. Martinez AR, Bernardus RE, Vermeiden JP, Schoemaker J. Reliability of home urinary LH tests for timing of insemination: a consumer's study. Hum Reprod 1992 Jul;7(6):751-753. [Medline: 1500469]
14. Behre HM, Kuhlage J, Gassner C, Sonntag B, Schem C, Schneider HP, et al. Prediction of ovulation by urinary hormone measurements with the home use ClearPlan Fertility Monitor: comparison with transvaginal ultrasound scans and serum hormone measurements. Hum Reprod 2000 Dec;15(12):2478-2482. [doi: 10.1093/humrep/15.12.2478] [Medline: 11098014]
15. Smoley BA, Robinson CM. Natural family planning. Am Fam Physician 2012 Nov 15;86(10):924-928 [FREE Full text] [Medline: 23157145]
16. Su H, Yi Y, Wei T, Chang T, Cheng C. Detection of ovulation, a review of currently available methods. Bioeng Transl Med 2017 Dec;2(3):238-246 [FREE Full text] [doi: 10.1002/btm2.10058] [Medline: 29313033]
17. Shilaih M, Goodale BM, Falco L, Kübler F, De Clerck V, Leeners B. Modern fertility awareness methods: wrist wearables capture the changes in temperature associated with the menstrual cycle. Biosci Rep 2018 Dec 21;38(6):1-12 [FREE Full text] [doi: 10.1042/BSR20171279] [Medline: 29175999]
18. Barron ML, Fehring RJ. Basal body temperature assessment: is it useful to couples seeking pregnancy? MCN Am J Matern Child Nurs 2005;30(5):290-6; quiz 297. [doi: 10.1097/00005721-200509000-00004] [Medline: 16132004]
19. Manhart MD, Duane M, Lind A, Sinai I, Golden-Tevald J. Fertility awareness-based methods of family planning: a review of effectiveness for avoiding pregnancy using SORT. Osteopath Fam Phys 2013 Jan;5(1):2-8. [doi: 10.1016/j.osfp.2012.09.002]

20. Berglund SE, Lindén HA, Scherwitzl R. Identification and prediction of the fertile window using NaturalCycles. Eur J Contracept Reprod Health Care 2015;20(5):403-408. [doi: 10.3109/13625187.2014.988210] [Medline: 25592280]
21. Berglund SE, Gemzell DK, Sellberg JA, Scherwitzl R. Fertility awareness-based mobile application for contraception. Eur J Contracept Reprod Health Care 2016 Jun;21(3):234-241 [FREE Full text] [doi: 10.3109/13625187.2016.1154143] [Medline: 27003381]
22. Epstein DA, Lee NB, Kang JH, Agapie E, Schroeder J, Pina LR, et al. Examining menstrual tracking to inform the design of personal informatics tools. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May 2;2017:6876-6888 [FREE Full text] [doi: 10.1145/3025453.3025635] [Medline: 28516176]
23. Sohda S, Suzuki K, Igari I. Relationship between the menstrual cycle and timing of ovulation revealed by new protocols: analysis of data from a self-tracking health app. J Med Internet Res 2017 Dec 27;19(11):e391 [FREE Full text] [doi: 10.2196/jmir.7468] [Medline: 29180346]
24. Shilaih M, Clerck VD, Falco L, Kübler F, Leeners B. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. Sci Rep 2017 Dec 2;7(1):1294 [FREE Full text] [doi: 10.1038/s41598-017-01433-9] [Medline: 28465583]
25. Baker FC, Waner JI, Vieira EF, Taylor SR, Driver HS, Mitchell D. Sleep and 24 hour body temperatures: a comparison in young men, naturally cycling women and women taking hormonal contraceptives. J Physiol 2001 Feb 1;530(Pt 3):565-574 [FREE Full text] [Medline: 11158285]
26. Moran VH, Leathard HL, Coley J. Cardiovascular functioning during the menstrual cycle. Clin Physiol 2000 Nov;20(6):496-504. [Medline: 11100398]
27. Sato N, Miyake S, Akatsu J, Kumashiro M. Power spectral analysis of heart rate variability in healthy young women during the normal menstrual cycle. Psychosom Med 1995;57(4):331-335. [Medline: 7480562]
28. Bai X, Li J, Zhou L, Li X. Influence of the menstrual cycle on nonlinear properties of heart rate variability in young women. Am J Physiol Heart Circ Physiol 2009 Aug;297(2):H765-H774 [FREE Full text] [doi: 10.1152/ajpheart.01283.2008] [Medline: 19465541]
29. de Zambotti M, Nicholas CL, Colrain IM, Trinder JA, Baker FC. Autonomic regulation across phases of the menstrual cycle and sleep stages in women with premenstrual syndrome and healthy controls. Psychoneuroendocrinology 2013 Nov;38(11):2618-2627 [FREE Full text] [doi: 10.1016/j.psyneuen.2013.06.005] [Medline: 23850226]
30. Tenan MS, Brothers RM, Tweedell AJ, Hackney AC, Griffin L. Changes in resting heart rate variability across the menstrual cycle. Psychophysiology 2014 Oct;51(10):996-1004. [doi: 10.1111/psyp.12250] [Medline: 24942292]
31. Hessemer V, Brück K. Influence of menstrual cycle on shivering, skin blood flow, and sweating responses measured at night. J Appl Physiol (1985) 1985 Dec;59(6):1902-1910. [doi: 10.1152/jappl.1985.59.6.1902] [Medline: 4077797]
32. Stachenfeld NS, Silva C, Keefe DL. Estrogen modifies the temperature effects of progesterone. J Appl Physiol (1985) 2000 May;88(5):1643-1649 [FREE Full text] [doi: 10.1152/jappl.2000.88.5.1643] [Medline: 10797125]
33. Eisenbeiss C, Welzel J, Schmeller W. The influence of female sex hormones on skin thickness: evaluation using 20 MHz sonography. Br J Dermatol 1998 Sep;139(3):462-467. [doi: 10.1046/j.1365-2133.1998.02410.x] [Medline: 9767291]
34. Tempdrop. How It Works URL: https://www.temp-drop.com/pages/how-it-works [accessed 2019-02-16] [WebCite Cache ID 76DSuS633]
35. YONO Labs. 2018. How YONO Detects the Fertility Window URL: https://www.yonolabs.com/how-yono-detects-the-fertility-window [accessed 2019-02-15] [WebCite Cache ID 76DS8Shto]
36. Kosecki D. FitBit. 2018. One of Your Most Requested Features is Here! Introducting Female Health Tracking URL: https://blog.fitbit.com/female-health-tracking/ [accessed 2019-02-20] [WebCite Cache ID 76KxHvQw1]
37. Apple Inc. 2019. A bold way to look at your health URL: https://www.apple.com/ios/health/ [accessed 2019-02-20] [WebCite Cache ID 76KxyZg6L]
38. Aguinis H, Gottfredson RK, Culpepper SA. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. J Manage 2013 Apr 2;39(6):1490-1528. [doi: 10.1177/0149206313478188]
39. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen 2014 Oct;143(5):2020-2045. [doi: 10.1037/xge0000014] [Medline: 25111580]
40. Koot P, Deurenberg P. Comparison of changes in energy expenditure and body temperatures after caffeine consumption. Ann Nutr Metab 1995;39(3):135-142. [doi: 10.1159/000177854] [Medline: 7486839]
41. Mannara G, Salvatori G, Pizzuti G. Ethyl alcohol induced skin temperature changes evaluated by thermography. Preliminary results. Boll Soc Ital Biol Sper 1993 Oct;69(10):587-594. [Medline: 8198799]
42. Tagliabue A, Terracina D, Cena H, Turconi G, Lanzola E, Montomoli C. Coffee induced thermogenesis and skin temperature. Int J Obes Relat Metab Disord 1994 Aug;18(8):537-541. [Medline: 7951473]
43. Torii M, Yamasaki M, Sasaki T, Nakayama H. Fall in skin temperature of exercising man. Br J Sports Med 1992 Mar;26(1):29-32 [FREE Full text] [Medline: 1600450]
44. O'Brien C. Reproducibility of the cold-induced vasodilation response in the human finger. J Appl Physiol 2005 Apr;98(4):1334-1340 [FREE Full text] [doi: 10.1152/japplphysiol.00859.2004] [Medline: 15579576]

45. Colantuoni C, Henry G, Zeger S, Pevsner J. Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts. Biotechniques 2002 Jun;32(6):1316-1320 [FREE Full text] [doi: 10.2144/02326mt02] [Medline: 12074162]
46. Nimon K. Statistical assumptions of substantive analyses across the general linear model: a mini-review. Front Psychol 2012;3:322 [FREE Full text] [doi: 10.3389/fpsyg.2012.00322] [Medline: 22973253]
47. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
48. Kuznetsova A, Brockhoff P, Christensen R. The Comprehensive R Archive Network. 2016. lmerTest: Tests in linear mixed effects models URL: https://cran.r-project.org/web/packages/lmerTest/index.html [accessed 2019-04-12] [WebCite Cache ID 77ZY39QBw]
49. Nash J, Varadhan R. Unifying optimization algorithms to aid software system users: optimx for R. J Stat Soft 2011;43(9):1-14 [FREE Full text] [doi: 10.18637/jss.v043.i09]
50. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J 2008 Jun;50(3):346-363. [doi: 10.1002/bimj.200810425] [Medline: 18481363]
51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011:2825-2830 [FREE Full text]
52. Alickovic E, Subasi A. Medical decision support system for diagnosis of heart arrhythmia using DWT and Random Forests classifier. J Med Syst 2016;40(108):1-12. [doi: 10.1007/s10916-016-0467-8] [Medline: 26922592]
53. Johnson S, Marriott L, Zinaman M. Can apps and calendar methods predict ovulation with accuracy? Curr Med Res Opin 2018;34(9):1587-1594. [Medline: 29749274]
54. Grrishma B, Gaur G, Velkumary S, Subramanian SK, Gurunandan U. Assessment of cardiovascular autonomic functions and baroreceptor reactivity in women with premenstrual syndrome. Indian J Physiol Pharmacol 2015;59(2):148-154. [Medline: 26685501]
55. Bartelink M, Wollersheim H, Theeuwes A, van Duren D, Thien T. Changes in skin blood flow during the menstrual cycle: the influence of the menstrual cycle on the peripheral circulation in healthy female volunteers. Clin Sci (Lond) 1990 May;78(5):527-532 [FREE Full text] [Medline: 2162282]
56. van Beek E, Houben A, van Es PN, Willekes C, Korten E, de Leeuw PW, et al. Peripheral haemodynamics and renal function in relation to the menstrual cycle. Clin Sci (Lond) 1996 Aug;91(2):163-168. [Medline: 8795439]
57. Huang S, Wong A, Ho CW, Weng TP, Cheng SC, Wang JS. Comparison of cardiac autonomic nervous system disturbed by sleep deprivation in sex and menstrual phase. Chin J Physiol 2015 Apr 30;58(2):114-123. [doi: 10.4077/CJP.2015.BAD287] [Medline: 25858472]
58. Robinson J, Ellis J. Mistiming of intercourse as a primary cause of failure to conceive: results of a survey on use of a home-use fertility monitor. Curr Med Res Opin 2007 Feb;23(2):301-306. [doi: 10.1185/030079906X162863] [Medline: 17288684]
59. Freundl G, Godehardt E, Kern P, Frank-Herrmann P, Koubenec H, Gnoth C. Estimated maximum failure rates of cycle monitors using daily conception probabilities in the menstrual cycle. Hum Reprod 2003 Dec;18(12):2628-2633. [doi: 10.1093/humrep/deg488] [Medline: 14645183]

Abbreviations

BBT: basal body temperature
BMI: body mass index
bpm: beats per minute
HRV: heart rate variability
LH: luteinizing hormone
NFP: natural family planning
OV: ovulation
WST: wrist skin temperature

Edited by G Eysenbach; submitted 15.01.19; peer-reviewed by A Benis, A Billis, M Boukhechba, L Lenzo; comments to author 06.02.19; revised version received 14.03.19; accepted 24.03.19; published 18.04.19.

Please cite as:
Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B
Wearable Sensors Reveal Menses-Driven Changes in Physiology and Enable Prediction of the Fertile Window: Observational Study
J Med Internet Res 2019;21(4):e13404
URL: http://www.jmir.org/2019/4/e13404/
doi:10.2196/13404
PMID:30998226

Review

Usability of Health Information Websites Designed for Adolescents: Systematic Review, Neurodevelopmental Model, and Design Brief

Gurpreet Kaur Reen[1,2], PhD; Linden Muirhead[3], PGD; Dawn Wendy Langdon[1], PhD

[1]Royal Holloway, University of London, Egham, United Kingdom
[2]University of Oxford, Oxford, United Kingdom
[3]MS Trust, Letchworth Garden City, United Kingdom

Corresponding Author:
Gurpreet Kaur Reen, PhD
University of Oxford
Department of Experimental Psychology
Oxford, OX2 6AE
United Kingdom
Phone: 44 7878451844
Email: Gurpreet.reen.2014@live.rhul.ac.uk

Abstract

Background: Adolescence is a unique developmental period characterized by biological, social, and cognitive changes, as well as an interest in managing one's own health care. Many adolescents use the internet to seek health care information. However, young people face barriers before they can understand and apply the health information that they access on the web. It is essential that usability of adolescent health websites on the internet is improved to help adolescents overcome these barriers and allow them to engage successfully with web-based health care content.

Objective: The aim of this review was to synthesize the usability of specific health information websites. These findings were mapped onto the adolescent neurodevelopmental profile, and a design brief based on the findings was developed to tailor future websites for specific adolescent requirements.

Methods: A systematic search conducted using PubMed, PsycINFO, and Education Resources Information Center (ERIC) identified 25 studies that assessed the usability of health information websites. Adolescent feedback was collected by a mixture of surveys, focus groups, interviews, and think-aloud procedures.

Results: A majority of the information websites were developed for specific health issues that may be relevant to adolescents. The most preferred website features were interactive content such as games and quizzes, as well as videos, images, audio clips, and animations. Participants also preferred communicating with other adolescents with similar conditions or learning about their experience through real stories and testimonials. Adolescents found it difficult to use health information websites if they contained too much text, were too cluttered, or had features that made it difficult to access. The findings are considered in the context of adolescent social processes, low tolerance of delayed gratification, and attraction to novelty and mapped onto a neurodevelopmental model of adolescence.

Conclusions: Young people's feedback can determine usability and content that make a health information website easy or informative to use. Neurodevelopmental profiles and the users' specific preferences and skills should be addressed in future development of health information websites for adolescents.

(J Med Internet Res 2019;21(4):e11584) doi:10.2196/11584

KEYWORDS

adolescents; health communication; internet; information seeking behavior; usability; systematic review

Introduction

Background

Adolescence is a significant period of development between childhood and adulthood (occurring approximately between the ages of 10 and 22 years) that is generally characterized by changes in social skills, cognitive capabilities, and a keen interest in one's own health care and well-being [1,2]. To better understand and manage their health, adolescents often look for information about specific health areas including sensitive topics such as sexual and mental health, as well as general guidance on health-related matters such as physical fitness and diet [3,4]. The internet is an appealing and valuable source for adolescents to seek this information [4-8]. A survey of over 1000 young people found that 84% of adolescents reported having used the internet at least once in their lifetime to access health information [3]. The advantages of the internet as a tool to find health information include easy access to a wide range of health topics, the ability to reach a large audience, and the ease with which adolescents can search for sensitive information anonymously [3,9]. Nevertheless, adolescents do not always find it easy to access the health information they find on the web.

Barriers to accessing health information often stem from low health literacy skills in the younger population. In the context of web-based information seeking, low health literacy in younger people can result in difficulty to either locate relevant health information on the internet, evaluate the credibility of the information they find, or apply this information to their personal life [4,6,7]. For instance, a large systematic review identified that adolescents find it difficult to create appropriate search strings when using search engines to find health information on the web [6,10]. Particularly, adolescents are likely to misspell medical terms resulting in a search that is irrelevant and confusing [7,10]. With regard to credibility, adolescents are keen to access information from reputable sources and generally avoid sites that are overtly commercial with excessive advertisements and links [6]. Furthermore, they may also stop their search as soon as they find the health information they need without comparing it with other possible sources [6,10]. Adolescents are thus not always equipped with appropriate skills and strategies to access credible web-based health information.

To overcome such barriers, accurate and evidence-based health information should be provided in a website customized for the younger population so that adolescents do not have to use unreliable internet sources [11]. In addition to content, the usability of these websites should be designed for the younger audience to ensure that adolescents are able to engage with the health care content [12]. For instance, a website that is poorly organized and presented can prevent young people from understanding and sometimes even retrieving the health information they require [10]. Improving the usability of websites with health information (referred now on as health information websites) can be a simple solution to the barriers faced by young people on the web. One method of improving usability is to incorporate the views and preferences of the users during the development of health information websites. By employing this user-centered approach, websites can be specifically tailored to the adolescent population, which can increase the likelihood that young people will access and regularly use the website content to better manage their health care [12-14].

Furthermore, it is now understood that adolescence is characterized by specific neurobiological, cognitive, and social development. Typical adolescent development includes, for example, a peak in sensation seeking and poor tolerance of delayed gratification, which are likely to affect how adolescents experience health information websites. The increased sensation seeking leads to risk taking, which has been explained by how differently the adolescent brain processes affective and social information, both likely to be found in health information websites [15,16].

Despite a large number of adolescent health information websites available, only a few have evaluated their usability with feedback from the younger population. A recent study reviewed general web-based health–searching behavior in adolescents [17]. However, there has yet to be a review that synthesizes adolescent feedback on the usability of specific health information websites.

Objectives

The purpose of this review was to determine adolescent preferences for the usability of specific health information websites and highlight the difficulties adolescents face when attempting to access the content on these sites. It is expected that this review will help inform the development of new adolescent health information websites, as well as allow existing adolescent health information websites to be adapted to their target audience. The findings of this review were considered in the context of the neurodevelopmental profile of adolescence including social processes, low tolerance of delayed gratification, and attraction to novelty.

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations were used as guidelines for the presentation of this review [18]. A protocol for this review has not been previously published or registered.

Systematic Literature Search

A systematic literature search was conducted in April 2018 using PubMed, PsycINFO, and Education Resources Information Center (ERIC). Uniform search terms were developed and used with all 3 databases (see Textbox 1).

Eligibility Criteria

Studies were included in this review if the following criteria were met: peer-reviewed original studies in English, studies with some adolescent participants aged between 13 and 17 years, studies about a health information website for any health topic, and studies that conducted some form of usability testing of the website from the perspective of adolescents. Both clinical and nonclinical adolescent populations were included into this review. Health information websites were defined as health websites which predominantly provide information about a specific health topic or provide general health guidance. The definition of health websites used in this review was adopted from the study by Sillence et al [19]. All types of study designs from the year 2000 were included if the inclusion criteria were met. The year 2000 was chosen so that a reasonable number of websites could be included in the review, although remaining moderately current.

Textbox 1. Search terms to shortlist studies.

adolescents OR young people OR teen OR young adult OR youth

AND

health OR nutrition OR sex OR illness OR medical condition OR disease OR alcohol OR tobacco OR physical activity OR diabetes OR cancer OR weight

AND

website OR online OR internet OR web 2.0

AND

evaluation OR views OR opinion OR survey OR usability OR feasibility OR user feedback

Studies were excluded if they only included participants aged older than 18 years or younger than 13 years. Studies that evaluated any other form of web-based health content were excluded (eg, health information presented through social media platforms, web-based courses, short-term education modules, support groups, and tools to collect patient data). Studies were also excluded if they evaluated usability of health websites that did not predominantly provide information about health topics. Studies were also excluded if they mostly focused on changing adolescent health behavior. It was expected that these studies evaluate usability features that are unique to websites designed to change health behavior but are not usually present in typical health information websites (eg, tools to monitor health behavior, feedback on progress, record of goals and plans, tailored content, review of previous health behavior, and messages of encouragement and motivation). Studies that only described a health information website, did not evaluate usability from the feedback of its users, or only described a protocol for a study were further excluded. Studies evaluating a nonhealth–related information website were also excluded from this review.

All titles and abstracts were screened. Studies that were deemed suitable from additional reference checking or searching through journals were also included. After initial screening, 302 studies were considered for eligibility and full texts were subsequently accessed.

Data Extraction

Data extraction forms were designed to extract relevant information from full texts and assess their eligibility for the final review. Extraction was initially carried out by 1 reviewer (GR) and was then verified by another (DL). Any discrepancies between extractions were resolved by discussion. After data extraction, 25 studies were included into the final review (see Figure 1).

All the studies were broadly qualitative, in the sense that they collected interview data and utilized ad hoc surveys. The narrative synthesis in this review was primarily based on the qualitative findings from these studies. Participant demographics were extracted from the shortlisted studies, comprising (where reported) age, gender, and whether participants were from the nonclinical or clinical population. Relevant information about the health information websites were also extracted, including the topic of the website and the characteristics of the website. The authors also noted whether the website was currently on the web. The method used to evaluate usability was also recorded. Any data that recorded adolescents' feedback on the usability of the website were extracted. Data were extracted from information in texts, tables, figures, and qualitative feedback.

Results were categorized into themes based on 5 of the 8 areas important for the development of health-related websites, as highlighted by Ritterband et al [12]. These areas are relevant to the usability of websites and include appearance (eg, visual appearance, organization of information, and screen size), burdens (eg, barriers to navigation), delivery of content (eg, animations, videos, illustrations, vignettes, and testimonials), message source (eg, credibility and age-appropriateness of the website), and participation (eg, the degree of interaction by the users). Understanding the adolescents' preferences for each of these domains could inform how each area should be tailored for the adolescent population when developing websites. The 3 remaining areas of website development were beyond the scope of this review (ie, content of the website, behavioral prescriptions or guidelines, and adapting website content per users' assessment and feedback) [12]. Results based on each of the 5 areas were further categorized by clinical or nonclinical adolescent users, by younger (aged ≤14 years) or older adolescents (aged >14 years), and by adolescents' gender.

Quality Rating

The Critical Appraisal Skills Programme (CASP) tool was used to assess the quality of the studies in this review by 2 independent reviewers (GR and DL). This tool was chosen as it has been recommended for reviewers when appraising qualitative studies [20] and has previously been used in other health-related systematic reviews [21,22].

The CASP tool is a checklist comprising 10 questions that broadly assess aims and methodology, appropriateness of research design, data collection, data analysis, findings, impact of investigator, ethics and values, and implications of research [20]. Reviewers are asked to rate each question on the checklist as yes, no, or unclear. Each of the questions are followed by prompts to help reviewers consider the criteria. The checklist also encourages reviewers to record the reasons for their response to each question. CASP is considered as a tool to assess the general quality of the study but does not offer any objective reason to exclude a study from the review. Therefore, no studies were excluded from this review based on the quality rating. There were only slight discrepancies between reviewers when rating study quality, which were resolved by discussion until an agreement was reached.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart for selection process of studies in systematic review. ERIC: Education Resources Information Center.

Results

Participant Demographics

A total of 25 studies evaluating usability of health information websites for adolescents were included into this review, and a quality appraisal was done for all studies (see Multimedia Appendix 1). A total of 2621 participants were recruited across all studies, ranging from ages 11 to 25 years (see Table 1). The majority of the website users were from a nonclinical population: 79.13% (2074/2621), followed by adolescents diagnosed with diabetes: 12.28% (322/2621), juvenile idiopathic arthritis: 2.06% (54/2621), hemophilia: 1.79% (47/2621), depression: 1.60% (42/2621), cancer: 0.84% (22/2621), cystic fibrosis: 0.61% (16/2621), migraines: 0.45% (12/2621), and those that had recently undergone a kidney transplant: 0.80% (21/2621). The mean age of participants was 15.2 years. Most studies included both younger (≤14 years) and older adolescents (>14 years).

In total, 9 studies did not provide enough information to calculate mean age [5,27,28,32,34,36,37,41,46], and 1 study only provided a median age [39]. Participants were recruited from a range of sources, including middle schools, secondary schools, specialist clinics, online, youth services, those who had participated in another study, and those who had already accessed the website. One study did not specify the location of recruitment [34].

Table 1. Participant demographics of included studies (n=25).

Study (first author, year)	Health status	Sample size (n)	Age range (years)	Age (years), mean (SD[a])	Gender	Recruited from
Ammerlaan, 2015 [23]	Diagnosed with juvenile idiopathic arthritis	13	17-22	20	92% females, 8% males	Outpatient clinics
Baulch, 2010 [24]	Nonclinical	67	Not specified	15.9 (1.8)	77% females, 23% males	Secondary school and a health obesity treatment program
Breakey, 2013 [25]	Diagnosed with hemophilia	18	13-18	15.5	Not specified	Hematology clinics
Breakey, 2014 [26]	Diagnosed with hemophilia	29	13-18	15.9	100% males	Hematology clinics
Coyne, 2016 [27]	Diagnosed with cystic fibrosis	16	15-25	Not specified	Not specified	Participants from previous study
Cullen, 2013 [28]	Nonclinical	211	12-17	Not specified	54% females, 46% males	Health fairs, schools, churches, community organizations, and newspaper and radio advertisements
Danielson, 2016 [29]	Nonclinical	18	13-18	16.3 (1.5)	100% females	Department of juvenile justice, high schools, youth centers, and snowball recruitment methods
Debar, 2009 [30]	Nonclinical	82	14-16	15.6 (0.6)	100% females	Participants from previous study
Donovan, 2012 [31]	Diagnosed with migraines	12	12-17	14	50% females, 50% males	Newspaper advertisements and community message board
Ercan, 2006 [32]	Nonclinical	105	13-15	Not specified	56% females, 44% males	Comprehensive school and special hospital school for children with depression and anxiety
Franck, 2007 [5]	Nonclinical	116	11-18	Not specified	Not specified	Secondary schools
Hanberger, 2013 [33]	Diagnosed with diabetes	287	Not specified	Intervention: 13.2 (3.7); Control: 13.3 (3.7)	Intervention: 52% females, 48% males; Control: 51% females, 49% males	2 pediatric clinics
Korus, 2015 [34]	Undergone kidney transplantation	21	12-17	Not specified	33% female, 67% males	Not specified
Long, 2009 [35]	Group 1: History of chronic pain; Group 2: Current diagnosis of chronic pain	Group 1: 5; Group 2: 6	Group 1: 13-17; Group 2: 12-16	Group 1: 15.8 (1.2); Group 2: 14.6 (1.6)	Group 1:60% females, 40% males; Group 2:67% females, 33% males	Multidisciplinary pediatric chronic pain clinic
McCarthy, 2012 [36]	Nonclinical	67	16-22; 72% aged between 16 and 17	Not specified	75% females, 25% males	Sexual health clinics, colleges, youth work contacts, youth websites, and friends of participants
Michaud, 2003 [37]	Nonclinical	Approximately 1394	13-18	Not specified	63% females, 37% males	Users of website
Nicholas, 2012 [38]	Diagnosed with diabetes	31	12-17	14.5	Not specified	Pediatric hospital
Nordfeldt, 2010 [39]	Diagnosed with diabetes	4	11-18	14[b]	Not specified	Clinical intervention study
Radovic, 2017 [40]	Diagnosed with depression	27 (Phase 1: 23; Phase 2: 4)	13-21	Phase 1: 16 (2.3); Phase 2: not specified	Phase 1: 78% females, 22% males; Phase 2 not specified	Phase 1: Academic adolescent medicine clinic and specialty psychiatric clinic; Phase 2: Youth advisory board
Radovic, 2018[41]	Diagnosed with depression	15	14-19	Not specified	Not specified	Clinical settings and online

Study (first author, year)	Health status	Sample size (n)	Age range (years)	Age (years), mean (SD[a])	Gender	Recruited from
Starling, 2015 [42]	Nonclinical	10	11-13	12.7	70% females, 30% males	Middle school
Stinson, 2015 [43]	Diagnosed with cancer	22	Not specified	15.2 (1.8)	41% females, 59% males	2 pediatric cancer treatment centers
Stinson, 2010a [44]	Diagnosed with juvenile idiopathic arthritis	19	Not specified	15.7 (1.5)	74% females, 26% males	2 rheumatology clinics in pediatric tertiary care centers
Stinson, 2010b [45]	Diagnosed with juvenile idiopathic arthritis	22	Not specified	14.4 (1.3)	68% females, 32% males	4 pediatric tertiary care centers
Wozney, 2015 [46]	Nonclinical	4	15-20	Not specified	50% females, 50% males	People involved in peer advocacy for mental health and illness

[a] Not all studies report SD with mean age.

[b] Median.

Website Characteristics

The majority of studies evaluated health information websites that were designed for specific health topics for adolescents (see Table 2). These include health information websites dedicated to weight management [24], physical activity and diet [28,30], organ transplants [34], transition from pediatric to adult health care [27], diabetes [33,38,39], juvenile idiopathic arthritis [23,44,45], human papillomavirus [42], anxiety [46], hemophilia [25,26], HIV prevention [29], depression [40,41], chronic pain [35], migraines [31], and cancer [43]. Health information websites that were developed to give general information to adolescents included a combination of health topics such as mental health, diet, drugs and alcohol, contraception, and sleep patterns. Furthermore, 4 health information websites were described by more than 1 study [25,26,33,39-41,44,45]. Some websites were already live on the internet before usability was fully evaluated [5,27,32,33,37,39], and several health information websites were still found to be on the web at the time of this review [23,25-27,29,34,37,40,41,44,45] (see Multimedia Appendix 2). One study did not provide the name of the health information website being evaluated [24].

User Feedback

Studies conducted with a clinical adolescent population used several different methods to collect usability feedback. In total, 2 studies used surveys to identify whether users were likely to use and accept the website features [26,31,35,41,45]. Furthermore, 4 studies employed a think-aloud procedure by recording users' thoughts as they used the website [25,27,34,40,44]. Studies with clinical adolescents also used interviews to record adolescents' experience of using the website [23,25,34,40,41,43,44]. One study collected information about the most visited website features [33], whereas another study requested an essay from its users about the best and worst website usability features [39].

With the exception of a think-aloud procedure, studies with nonclinical adolescent populations also collected usability feedback from users through surveys [5,24,28-30,32,37], website visits [30,37], and interviews [5,29]. One study also conducted focus groups with nonclinical adolescent users before developing the website [36].

Table 2. Website characteristics and patient feedback about website usability (n=25).

Study (first author, year)	Topic of website	Website features	Method of evaluation	Positive feedback	Negative feedback
Ammerlaan, 2015 [23]	Juvenile Idiopathic Arthritis	Log-in feature, communication with a specialist, access to medical record, self-monitoring (eg, online diary)	Interviews	Preferred features: Website design, easy to read, well-targeted to young adults, videos, life stories of other patients. Features wanted: Facility to make online appointments, access to x-rays, printing forms for blood collection, opening hours of specialists. Most used tool: Access to personal medical records	Features not preferred: Log-in code too long and complex, access to medical records overwhelming, Least used tool: Communication with specialist (n=4) Self-monitoring (n=4)
Baulch, 2010 [24]	Weight management	Monitoring, reward ideas, links to additional ideas, charts to plot daily activity, color, interactive, discussion board, health professional to answer questions	Modified Technology Acceptance Model scale (12-item measure of usefulness and ease of use on 1-7 Likert scale) Survey	Perceived usefulness: (M=5.04, SD=0.98), perceived ease of use (M=5.48, SD=0.98), intention to use (M=4.82, SD=1.48). Most preferred features: Links to additional ideas (97%), charts (84%), email feedback (82%), login (79%), links to other websites (76%), phone support (61%), discussion board (57%), chat room (52%). Additional preferred features: Games, success stories, motivation quotes	Least preferred feature: Online journal (37%)
Breakey, 2013 [25]	Hemophilia	80 web pages of content, images, interactive animations, quizzes, glossary, self-management strategies	Think-aloud procedure; Interviews	Preferred features: Color, chunking of information, images, animations, interview of peers with condition	Features not preferred: Videos too dark (n=3), Venn diagrams difficult to comprehend
Breakey, 2014 [26]	Hemophilia	80 web pages of content, images, interactive animations, quizzes, glossary, self-management strategies	Survey	Preferred features: Easy to use, videos, animations, relaxation exercises	Least used feature: Internet forum
Coyne, 2016 [27]	Transition from pediatric to adult healthcare	Home page, top-tips section, 9 video testimonials, FAQs, essential reading, external links, photo gallery	Think-aloud procedure	Preferred features: Good layout, easy navigation, testimonials, short length videos (2-5 min), age-appropriate, colorful Features wanted: Downloadable information sheets, visual images of clinics and hospitals	Features not preferred: Pages too long, too much information, bright orange color, picture of young person considered cheesy, did not understand meaning of FAQ
Cullen, 2013 [28]	Physical activity and diet	Log-in feature, healthy eating calculator, goal setting, 12 short role-model video stories, Did you know section, blog, track goal progress (diary), print goal sheet	Survey	Preferred features: Healthy eating calculator (89%), Goal setting (91%), Did you know section (88%)	Least used features: Diary used 3 or more times (33%)
Danielson, 2016 [29]	HIV prevention	Subsections, videos, photos of women, interactive quizzes	Survey; Interviews	Preferred features: Videos, age-appropriate, interactive activities and quizzes, pictures of other females	Features not preferred: Videos too long, mixed reviews on navigation (some found it easy to navigate and some found it difficult to navigate)

Study (first author, year)	Topic of website	Website features	Method of evaluation	Positive feedback	Negative feedback
Debar, 2009 [30]	Physical activity and diet	Log-in feature, bulletin board to communicate with peers and staff, special form to ask questions to staff for confidential reply, forum to post anonymous questions, handouts, links to external content, hot tips section providing short summaries, photos of staff, quizzes, options to win prizes, incentive points, My progress page	Survey; Website visits	Most preferred features: Incentive points system (50%), Learning new information (37%). Most visited pages: Fun stuff such as quizzes (35%), Social networking features (29%), Scrapbook (24%)	Least preferred features: Information not regularly updated (30%). Least visited pages: Hot tips summaries (8%), additional resources (4%)
Donovan, 2012 [31]	Migraines	Quizzes, audio and video-based tools, social networking, virtual toolbox of coping strategies, headache diary	Survey	Features preferred: Ask an expert feature, video-based content. Features wanted: Library of content	—[a]
Ercan, 2006 [32]	General Includes: mental health, eating problems, drugs and alcohol	interactive stories, cartoons about depression, games	Survey	Preferred features: Graphics/pictures (86.6%), site customization (85.1%), games (76.9%), interactive stories (83.1%)	—
Franck, 2007 [5]	General; Includes: body, chronic conditions, disabilities, tests, treatments	Sections dedicated to children, teens and families section	Analysis of 30 min website navigation; Survey; Informal discussion	Navigation: 45% went to older adolescent section before going to section suitable for age Preferred features: Pictures, games and animations, real stories. Features wanted: Positive recovery stories	Features not preferred: Gender imbalanced pages, too much text in children section, the cartoon characters used, real stories section too negative, older adolescent page considered dull and boring, disagreement for pages too crowded or too plain
Hanberger, 2013 [33]	Diabetes	Social networking, discussion board, local practitioners' details, information about local activities, new research, questions and answers, photos of staff, education videos, date of last update and people who wrote webpage	Website visits	Most visited pages: Home page (10%), stories section (5.3%), blogs (7.5%), questions and answers (1.8%), videos (2.7%), discussion board (1.2%)	Least visited page: External links (0%)
Korus, 2015 [34]	Transplant	Videos, colorful	Think-aloud procedure; Semistructured interviews	Preferred features: Visually-appealing with lots of color, video testimonials Features wanted: Hyperlinks to other pages within module, a search box, drop-down menus, more color, more pictures and graphics, a discussion forum, music, interactive qualities	Features not preferred: Mixed reviews about navigation (some found it easy to navigate and others found it difficult to navigate)
Long, 2009 [35]	Chronic pain	Video interviews of peers with condition, relaxation audio clips, 200 content pages with graphics, log-in, goal setting, interactive, questions and answers	Stage 1: Survey for users with past history of chronic pain (1-5 Likert scale). Stage 2: Survey for users with current chronic pain; 1-5 Likert scale)	Stage 1: Perceived ease of use: (M=4.40, SD=0.55). Stage 2 Perceived ease of use: (Mean=4.50, SD=0.64); Preferred features: Easy to navigate, video, audio, personalization	Stage 1: Features not preferred: Difficult to understand images, lengthy content. Stage 2: Features not preferred: Some images, long load times

Study (first author, year)	Topic of website	Website features	Method of evaluation	Positive feedback	Negative feedback
McCarthy, 2012 [36]	Sexual health	Interactive quizzes, short interactive activities, peer-to-peer exchange of views (website deigned after user feedback)	Focus groups	Preferred features: Social interaction, anonymous, videos about peers discussing real stories, commenting on videos, dramatic story format, images of people, images of scenarios, images about specific sexual health issues, easy to understand website logo, clear and memorable website name, weekly update, interactive activities	Features not preferred: Too much text, disagreement on color (some users preferred bold colors and others preferred neutral tones)
Michaud, 2003 [37]	General; Includes: drug, alcohol, contraception, sleep patterns	Emailed questions answered by professionals, addresses of professional institutions, list of previously asked questions	Survey; Website visits	Preferred features: Structured answers to questions, list of other questions, addresses of other institutions. Most visited page: Questions and answers (82%)	—
Nicholas, 2012 [38]	Diabetes	Not specified	Interviews	Preferred features: Discussion topics on forum, anonymity Features wanted: Personalization, instant-messaging	—
Nordfeldt, 2010 [39]	Diabetes	Social networking, discussion board, local practitioners' details, information about local activities, new research, questions and answers, photos of staff, educational videos, date of last update and people who wrote webpage	Qualitative essays	Preferred features: Facts, simple layout, easy to use and log-in	Features not preferred: Difficult chatting feature, difficult log-in as password too hard to recover
Radovic, 2017 [40]	Depression	Blog posts, Links to external resources, personalization, questions posed for user discussion	Phase 1: interviews Phase 2: Think-aloud procedure and System Usability Scale (1-5 Likert scale)	Phase 1: Most preferred features: online peer support, blogs, moderator to avoid sharing incorrect information with other peers, anonymity. Phase 2: 100% completed all think-aloud tasks; Usability score mean=4.5 (SD=.31)	Phase 2: Least preferred feature: Wiki tool to contribute to a story
Radovic, 2018 [41]	Depression	Blog posts, links to external resources, personalization, questions posed for user discussion	System usability scale (1-5 Likert scale); Survey; Interview	Preferred features: Age appropriate format, positive atmosphere, interactive, regular new content, positive stories, anonymity. Features wanted: Social interaction, interactive. Rating of user friendliness: 50% users thought user friendly of site was good	Features not preferred: Structured communication on forum, log-in
Starling, 2015 [42]	Human papillomavirus	Interactive quiz show game, texting stimulation, FAQ section	Analysis of navigation System usability scale (Rating of user friendliness, 1-7 Likert scale); Survey	Preferred features: Easy to use, website functions. Rating of user friendliness: (M=6.0). Features wanted: Clear title	Features not preferred: Sound levels of videos too high

Study (first author, year)	Topic of website	Website features	Method of evaluation	Positive feedback	Negative feedback
Stinson, 2010a [44]	Juvenile Idiopathic Arthritis	310 content pages, animations, images, videos of peers with medical condition, written stories, discussion boards, surveys, quizzes, glossary of medical terms, relaxation audio clips, visual imagery audio clips, printable PDF information forms for teachers, journal for symptom tracking and weekly goals, ask the expert features, personalization	Analysis of 30-45 min navigation; Think-aloud Semistructured interviews	Preferred features: Animations, chunking of text, layout, simple website, up-to-date, age-appropriate, glossary of medical terms, ask the expert, pdf forms, audio clips, journal, discussion boards. Features wanted: Chunking of texts using graphics and animations, labels on medical diagrams, important information on top of page	Navigation errors: Overall navigation errors (10%), Presentation error (26%), Control usage of animations and videos error (42%). Features not preferred: Teenager looking sad, control buttons on animations and videos not easy to see
Stinson, 2010b [45]	Juvenile Idiopathic Arthritis	310 content pages, animations, images, videos of peer with medical condition, written stories, discussion boards, surveys, quizzes, glossary of medical terms, relaxation audio clips, visual imagery audio clips, printable PDF information forms for teachers, journal for symptom tracking and weekly goals, ask the expert features, personalization	Survey	Preferred features: Easy to use, video, relaxation audio guide, visual imagery audio guide, personalization	—
Stinson, 2015 [43]	Cancer	200 contents pages, animations, images, videos, discussion boards, surveys and interactive forms, log-in required	Analysis of 30-40 min navigation; Think-aloud Semistructured interviews	Preferred features: Color scheme, sections and subsections making it easy to locate information, bright colors, graffiti-style wallpaper, simple words with not too much jargon, videos from experts, glossary of terms, interactive components, video clips of other adolescents. Features wanted: Font and appearance to take up more room, search bar at the top of the page	Features not preferred: Too much blank 'white' space
Wozney, 2015 [46]	Anxiety	Personalization (Email reminders, prepopulating answers, notifications), videos, log-in, check-in & check-out, homework activities	Interviews	Preferred features: Introductory animation video, personalization, comic book style, videos, images, graphics, animations, logos. Features wanted: More videos and images needed	Features not preferred: Slow page loading, no chunking, not visually pleasing

[a] Not applicable.

Visual Appearance

Websites that used bright colors were appreciated by adolescents with medical conditions [25,27,34,43], with users of 2 websites stating that they would like more color [27,34]. Color was preferred by both the younger [25,34] and older adolescent population [25,27,34,43], as well as by both genders [34,43]. One study did not evaluate the specific visual features of a website, but adolescents reported that they favored the overall appearance and design of the website [23]. Adolescent users also stated that they did not like a lot of blank space on the website's page [43] and preferred a website with a simple layout [44]. These users were mostly older male and female adolescents.

Similar to the clinical population, nonclinical adolescents did not like websites that looked too dull or boring [5] or had a lot of blank space on the page [5]. Websites with excessive text were also not considered visually appealing [5,36,46] and in fact led to users in 1 study to state that they did not like the overall visual appearance and website design [46]. This was true for both male and female adolescent groups [36,46], as well as younger and older adolescents [5,36,46].

Navigation Burden

A total of 10 studies recorded adolescents' navigation of health information websites. Adolescents diagnosed with medical conditions across 6 studies reported that they were easily able to navigate the website [27,35,39,40,44,45]. Even younger adolescent groups found the website easy to navigate [39,40], although a difficult log-in procedure was not appreciated [39]. In 1 study, users stated that they did not like the log-in procedure at all [41]. Users in another study offered mixed opinions about the navigation of the website; some users found the website easy to navigate, whereas other users found the website difficult to navigate [34]. These groups of users did not differ in age or gender compared with adolescents who felt generally positive about website navigation. With regard to improving navigation, adolescents suggested that websites should have a search bar [34,43], drop-down menus [34], information at the top of the page [44], and hyperlinks within the text, which link to other pages on the website [34].

Similar to the clinical population, nonclinical adolescents in 2 studies found websites easy to navigate [24,42], and users in 1 study differed in their opinions regarding navigation [29]. The only difference between users who found navigation easy and those with mixed opinions was that the latter group of adolescents were all females [29]. Specifically, users did not appreciate slow loading pages or a difficult log-in procedure when navigating the website [23,39]. In 1 study, users reporting difficulties logging in to the website actually used the website less compared with users who did not report these difficulties [23]. Adolescents in these studies were of all ages and mostly female [23,39].

Delivery of Content

Health content was delivered in a variety of ways across websites. Participants diagnosed with medical conditions strongly expressed their preference for content delivered using videos [23,26,27,31,33-35,43,45], images [25,34], audio clips [35,44,45], and animations [25,26,44]. However, adolescents also wanted videos to be clearly visible [26], controls on videos and animations to be clearly displayed [44], “cheesy” images to be removed [27], and all content delivery to be accessible and easy to comprehend [25,43]. In fact, users did not like images that were too difficult to understand [35] and stated that medical images on the website should have labels to improve clarity [44]. Adolescents also stated that they wanted to see more images of the clinics and hospitals that they will be attending [27]. Adolescents of both genders and all ages appreciated content delivery in this manner. They also strongly preferred information that was delivered through their peers. Adolescents really liked seeing real stories and testimonials from other adolescents who had the same health issues [25,27,34,41,43], especially when presented in a video format [27,34,43], and if the stories were positive in nature [41]. Adolescents also liked content that was generated by other users, such as blogs [33,40]. Other website features that delivered content and were preferred by clinical adolescent users included graffiti-style wallpaper [43], personalization of the website [35], subsections and chunking of text [25,43,44], glossary of medical terms [44], large number of facts [39], and relaxation exercises [26]. When some of these changes were made to 1 website, adolescents reported that they found the website much more appealing and expressed their intention to use the website in the future [25]. Despite preferring a range of features, adolescents also wanted websites to be simple and not too cluttered [39,43,44]. All clinical adolescent users in these studies were mixed in terms of gender and age range.

Similar patterns were observed in studies with nonclinical populations. Adolescents also expressed their preference for website content delivered using videos [29,36,46], animations [5], and images [5,24,29,32,36,46]. Specifically, images that presented graphs, charts, or logos; images of other adolescents; or images about the health issue were appreciated by users [24,29,32,36,46]. However, adolescents did not like videos that were too long [42], videos where sound could not be easily controlled [42], images that showed unlikeable cartoon characters [5], or graphics that were too difficult to comprehend [46]. Similar to the clinical adolescent group, nonclinical adolescent users also liked real stories and testimonials from their peers and from those with similar health issues [5,32,36,37]. However, adolescents also stated that although the real stories section was a good feature, these stories and testimonials needed to be more positive and show adolescents successfully managing their health issues [5]. Other website features that were preferred by nonclinical adolescent users included information about other institutions [37], links to other websites and ideas [24], clear logo and website name [36,46], and a comic book style [46]. All nonclinical adolescent users in these studies were also mixed in terms of gender and age range.

Message Source

The message source of the website and how this was perceived was explored in only a few studies in this review. Clinical adolescents in 3 studies reported that they found the website age-appropriate [27,41,44]. However, adolescents in these studies were all aged 14 years or above, and it is not possible to determine whether younger adolescents would also find the website to be age-appropriate.

Nonclinical adolescent users in 1 study also considered the website to be age-appropriate [29]. Adolescents in this study were aged between 13 and 18 years, indicating that the website was suitable for all adolescent age groups [29]. Other users also offered suggestions to improve the message source of websites. Adolescent users of 2 websites felt that the logo and website name should be clearer and more suitable to the target audience [36,46]. Adolescents in these studies were mixed in gender but were all in the older adolescent group. In 1 study, adolescents also noted that the webpages were somewhat gender imbalanced (ie, too “girly” or too much information about what boys like) and needed to be adapted for all its users [5]. However, this study did not specify the number of males and females that were evaluated.

Participation

Many websites in this review used interactive features that were generally favored by users of these websites. Specific interactive features that were preferred by adolescents with a clinical condition include games and quizzes [25,43], a journal feature [44], and the ability to personalize the website [44]. These adolescents ranged from the young to old and included both males and females [25,43]. One study added more interactive features to the website following usability testing, and users expressed their intention to use the website in the future following the changes [25]. Only 1 website let its older clinical users access personal medical records, a feature that was preferred and used regularly, although some found this too overwhelming [23]. Participants also appreciated health information websites that allowed them to interact with health professionals and their peers. For instance, clinically diagnosed adolescents liked it when health professionals provided answers to the questions that they had posed, particularly when the answers were structured in an accessible manner [31,44] or delivered in a video format [43]. However, communicating with a health professional was the least used website feature in 1 study [23]. Social networking with other peers was another interactive feature favored by website users in 3 studies [33,38,40]. These included social networking using online peer forums [40] and discussion boards [33,38,44]. In 1 study, adolescents stated that they wanted more social interaction in the website [41]. All these studies included adolescents across different ages and genders. Clinical adolescents in only 1 study did not use the social networking feature [26]. These adolescents were all male and ranged from the young to old. However, the authors noted that this was likely because adolescents could not navigate the social networking feature and it did not reflect the adolescents' preference for the feature itself [26]. Despite the general preference for social networking, adolescents with medical conditions wanted anonymity during social interaction [38,41] and did not show much inclination for text information that they could edit, such as wiki tools or the online journal feature on websites [40]. Participants also disliked if it was too difficult to participate in the interactive features of the website such as if the chatting feature was too difficult to use [39]. This negative feedback was not specific to a particular age range or gender.

The nonclinical adolescent population also expressed their preferences for more interactive features. These include games and quizzes [5,24,29,30,32,36,42], ability to set goals [28], ability to customize the website [32,46], a scrapbook feature [30], an incentive point-based system [30], and an interactive demonstration of health information [29]. Adolescents in these studies were generally mixed in age and gender. Similar to the clinical population, nonclinical adolescents preferred to interact with health professionals and their peers. For instance, participants liked it when health professionals provided structured and accessible answers to the questions that they had posed [37]. Social networking with other peers was another popular feature [24,30,36]. This includes social networking using discussion boards [24], chat rooms [24], and phone support [24]. Nonclinical adolescent users across these websites were of all ages and gender. One study found that participants did not like to use the online journal feature on the website [28]. This study did not differ in terms of users' age or gender compared with studies that only had positive feedback about the social networking feature.

Discussion

Principal Findings

With the advent of health information websites available on the internet, adolescents have access to a large array of content to manage their health and well-being [1,3]. However, adolescents face many challenges when obtaining and appraising web-based health information [6,10,47]. To overcome these barriers, it is essential that health information websites are suitably targeted to the adolescent population. This not only means providing relevant health-related content to young people but also improving the usability of websites so that young people are better able to access web-based health care information [12]. This systematic review was the first to synthesize the preferences and difficulties that adolescents face when using specific health information websites. The findings of this review can help tailor new and existing health care websites designed for adolescents and increase the likelihood that adolescents will successfully be able to access the health care information they require to better manage their health.

A total of 25 health care websites were identified by this review, the majority of which were designed to provide information about specific health topics. Health care content on these websites were mostly delivered using interactive features (eg, quizzes and games), videos, audio clips, animations and illustrations, and with some form of social networking. All these features were preferred by adolescents engaging with these websites, irrespective of medical diagnosis, age, and gender [5,23-32,34-36,38,40-46]. These features should be seriously considered when developing future health information websites for younger people. For instance, using interactive and digital content such as videos, images, and graphics, may help to chunk the text on a website's page, as suggested by the users of 1 website [44]. This is especially important as adolescents do not appreciate information presented in a continuous stream of text [5,35,36,46]. Nevertheless, any features presented along with the text should still serve a purpose and not make the webpage too cluttered, as this can prevent adolescents from browsing the website altogether [10]. Illustrations, videos, and animations should also be easy to access and understand to prevent adolescents from getting distracted [25,35,44]. Another commonly preferred feature on health care websites was stories and testimonials from other young people sharing their experience with the health issue [5,25,27,32,34,36,41,43,44]. Adolescents felt that they were able to resonate with other people's experiences more so than general facts about the health topic but preferred the experiences to be positive in nature. Adding this feature may also help to improve adolescents' motivation to continue engaging with the website, perhaps because they identify with the health issue presented in the story [12]. Future health care websites may consider combining testimonials and stories with other favored features to improve usability of web-based health care information for this population.

The findings from this review support the view that an archetypical “digital native” with an easy, intuitive facility for media may not be broadly representative of younger people [48] and that there is still a requirement in this group for improved usability of websites [6,10]. In addition, although young people (in the United States at least) spend 7.5 hours per day every day of the week engaged with media, one-third of this time involves juggling several media simultaneously. Multitasking disrupts concurrent learning (both in the classroom and during homework) and influences cognitive function on single tasks. This is even true of background television [49]. Despite this, health information websites are usually designed with the adolescent users' undivided attention in mind.

Adolescence is also a significant and unique period of development. The cognitive and socioaffective developments in adolescence are linked to significant dynamic changes in the brain structure and function [2,50], and the trajectories of structural development are remarkably consistent [51]. It is interesting to examine how far the feedback of adolescents regarding health information websites reflects their neurodevelopmental stage. White matter connections increase during this period, facilitating communication within the cerebral cortex. These connections eventually evolve to adult patterns, which is linked to growing behavioral control and better ability to wait for a reward (prefrontal connections to the subcortical striatum) [52]. These connections not yet being fully formed in adolescents could explain why adolescents have high intolerance of slow loading pages [35,41,46] and are likely to stop searching a website if they are unable to retrieve information quickly [10]. The feedback that real stories from other young people needed to be more positive and real stories should feature other adolescents who had successfully managed their health condition [5,41,44] could be linked to adolescents' increased sensitivity to happy faces compared with adults [53] who presumably designed the website. Brain regions that relate to social function (medial prefrontal cortex, superior temporal cortex, and temporal parietal junction) evolve quickly in adolescence [54,55], which may also explain the preference for social networking [24,30,33,36,38,40,41,44]. Nevertheless, no health information websites explicitly profess to have tailored their design for the cognitive skill profile of adolescents. However, some models of health literacy consider children's age-specific cognitive development [8], and there has been some consideration of how media use more generally relates to adolescent neural development [56]. Neurocognitive studies may be a way of assessing the efficacy of health messages delivered to the younger population [57]. Taking account of these and other related findings, we have constructed a neurodevelopmental model and corresponding design brief for adolescent health information websites (Table 3).

Table 3. Mapping of adolescents' feedback on health information websites, their neurodevelopmental underpinnings, and design indicators.

Adolescents' feedback	Specifics	Psychological profile	Neurodevelopmental underpinnings	Recommendations from adolescents	Design indicators
Visual appearance	Bright colors ✓[a]; Blank space x[b]; Too much text x	Novelty-seeking, high susceptibility to rewards [57]	Hypersensitivity of reward regions. There is an adolescent peak in reward-related ventral striatum, specifically nucleus accumbens, activity [58]	More color	Balance of white space to avoid clutter; clear correlation between input and output (reward value); consider accessibility of color options—target Web Content Accessibility Guidelines 2.0 conformance level AA; use images only where they add value; chunking pages into smaller bite-size bits
Navigation burden	Search features ✓; Use of hyperlinks ✓; Slow loading pages x; Difficult log-in x	Low tolerance for delayed gratification [59]	Delay tolerance positively correlates with the resting state functional connectivity between the dorsal anterior cingulate cortex and the left dorsolateral prefrontal cortex, a critical functional circuit in the cognitive control network [59]	Search bar; drop-down menus; hyperlinks to other pages	Consider weight of page, especially as images, graphics, and videos can add to slow loading; resize images to decrease weight on page; enable different ways to search and access, including effective tagging, use of synonyms, filtered and faceted search, and drop-down filters; create different ways for users to search or browse, including effective use of breadcrumbs and anchors; use language appropriate for adolescents for navigational menus; create user journeys with relevant links between sections of the site; use meaningful words to describe links; clearly display whether an in-site link or to an external site; user test log-in process; only capture data as required for fulfilment of service; clear links to privacy statements; clear wording around opt-in and opt-out permissions; 1 click log-in functions, that is, login through Google or Facebook
Delivery of content	Videos ✓; Images, graphics, and charts ✓; Audio ✓; Animation ✓	Processing narratives, emotional stimuli, self-relevance, and attention from salient stimuli [60]	Effective video health messages for young people involve responses in dorsomedial prefrontal cortex, insulae, and precuneus [61]	Sound easy to control; clearly visible; videos not too long; images or graphics not difficult to comprehend; websites that were not too cluttered	Position on page relevant to content; embed videos with uploaded image; play through YouTube, Vimeo, or dedicated video service; use of playlists to manage video length; ensure transcript available for accessibility; resize images; include meaningful alt tags; use images that add value (eg, infographics); consider use of suitable images to break up text, but must be relevant and not just decorative (especially as any image will add to page load); review how screen readers will deal with images
Delivery of content	Information from health professional ✓; Questions from adolescents ✓; Vignettes and testimonials from adolescents ✓	Conscious viewing of credentials for medical expertise and to judge credibility of web-based health information [62]. Especially sensitive to how adolescents fit into their social environment [63]	Neural substrates for social behavior overlap with neural substrates underlying physical pain. This involves the dorsal anterior cingulate, subgenual/ventral anterior cingulate cortex, right ventrolateral prefrontal cortex, medial prefrontal cortex, posterior cingulate, and insula. These regions overlap, continuing to develop structurally and functionally during adolescence [62]	Structured answers from health professionals; positive testimonials from other adolescents	Clear acknowledgment of expertise; ensure consent for any personal data disclosed; framing of adolescent testimonials

Adolescents' feedback	Specifics	Psychological profile	Neurodevelopmental underpinnings	Recommendations from adolescents	Design indicators
Message source	Gender imbalanced pages x; unclear logo and website name x; Age-appropriate ✓	Low tolerance for delayed gratification [59]	Delay tolerance positively correlates with the resting state functional connectivity between the dorsal anterior cingulate cortex and the left dorsolateral prefrontal cortex, a critical functional circuit in the cognitive control network [59]	Make websites age-appropriate; adapt websites to both gender	Use age-appropriate language; use of age-appropriate skins; ensure design reflects trends in adolescent behavior; recognize different age ranges with age-appropriate backgrounds, images, and personalization; consider age-based content segmentation; use colors and designs appropriate for all genders; clear positioning of website name and logo to create credibility in the brand
Participation	Quizzes ✓; Games ✓; Customized webpages ✓; Social networking ✓	Heightened attraction to novel and exciting experiences known as sensation seeking [57]; Friends are more important than at any other stage of life [64]	Rising dopaminergic activation during adolescence, which may reflect activity in the midbrain dopamine pathway ascending from the ventral tegmental region [57]	Interactive content	Break up text with interactive content; use games and quizzes to help teach health content, but must be relevant and necessary (especially as any interactive content will add to page load); polls to receive feedback from adolescents (such as polls on Twitter); allow personalization of webpages; enable peer-to-peer interaction and user-generated content; moderation to prevent trolling; manage safeguarding concerns; manage data protection concerns, access, and verification controls

[a] ✓: indicates that this is a good feature.

[b] x: indicates that this is not a good feature.

In addition to neurodevelopmental changes, many adolescent clinical groups have particular cognitive profiles associated with their diagnosis, and health information websites should attempt to tailor to these. For example, adolescents with autism have specific challenges processing emotional and social stimuli [58] and sensory reactivity [59], which develops over their lifetime. Evaluation of health information websites for people with autism and other intellectual disabilities has tended to focus on content rather than on the format and user interface [65,66]. Similarly, young people with multiple sclerosis can encounter diminished attention, processing speed, visuomotor skills, language, and general intellectual level, with a mixed picture of loss of previously acquired function and failure of continued progress to expected adult intellectual competencies [60]. Guidance for website design already exists for other clinical and carer groups (eg, [67-69]). It is important that health information websites address users' particular needs at the design and testing stage, with suitable user group engagement.

The feedback from young users of health information websites can be mapped onto their cognitive characteristics and associated neurocognitive substrates (Table 3). This has particular and specific implications for the technical and content aspects of website design. In fact, it is likely that the content and format of a website may compete, which is also known as argument quality and message sensation value [57,61]. Salient elements of a message may attract attention to the extent of reducing adolescents' available processing capacity to comprehend the message and remember it and its persuasiveness. It seems possible that this could be explained by a fundamental competition between emotion and cognition for neural resources because they share some neural networks [62]. This is particularly true if negative emotional images are followed by increasing cognitive demand, a sequence likely to occur in health information websites. This poses a dilemma for designers, which can only be resolved by focus groups and other usability evaluations with the appropriate client group.

In addition, the content of adolescent health information websites and how it is delivered could be tailored to connect more successfully with adolescents. Changes in processing of affective and social information make adolescents more likely to attend to short-term rewarding outcomes, whereas many health behavior decisions require consideration of long-term negative outcomes [15]. At the time of adolescence, meta-cognitive processes undergo major changes, leading to improved cognitive control in adulthood, probably as a result of maturation of association cortices (such as the prefrontal cortex, posterior parietal cortex, and superior temporal cortex [16]). The adolescent period of instability in working memory, response inhibition, and performance monitoring could be addressed in website design to support and accommodate these characteristics and ensure that health information is successfully and comfortably communicated.

It is interesting to compare our findings with health website feedback from other groups. Parents of children with attention-deficit/hyperactivity disorder (ADHD) reported on an educational website and some commented that the appearance could be more appropriate for individuals with disability and that the content could have been more suitable and engaging for adolescents with ADHD [70]. They suggested that the website was “boring”, or disengaging, especially for adolescents. Others suggested introducing games for younger children. The comments from parents relating to this adolescent population would seem to fit with the findings from this review. Furthermore, Scanlan et al [71] investigated a large group of healthy young adults (aged between 18 and 25 years, primarily female university students), slightly older on average than the population in this review. Each participant was given a vignette and asked to imagine that they had just received a depression diagnosis. They were randomized to view either a traditional health information website for all ages, which listed evidence-based depression treatment recommendations, or a website targeting those aged between 12 and 30 years, which aggregated reports from young people who lived with depression. Website design was rated to be significantly better on the site that aggregated service user accounts, but interestingly this was not related to behavioral intentions in this nonclinical sample. Treatment decisions were related to content that was perceived to be more credible and endorsed by experts, rather than other users’ accounts. This study showed that preference for website design was apparently not the main behavior driver for an older, healthy cognitively intact sample imagining they were depressed. Content credibility was a significant influence [71]. Therefore, the effect of preference for website design on behavior needs to be investigated in a younger adolescent group, for whom the content is personally relevant.

Strengths and Limitations

A recent review also examined adolescents’ preferences for web-based health information, as well as the type of health content that adolescents seek on the web [17]. However, Park et al did not review specific health information websites but rather focused on the adolescent’s general health-seeking behavior on the web [17]. Only 1 study in this review was the same as that included by Park et al [17]. This review extends these findings of Park et al by focusing on tangible usability features of specific adolescent health information websites based on feedback from its users (eg, the type of multimedia that should be used), leading to specific recommendations. Furthermore, this study mapped the findings of the systematic review onto a neurodevelopmental model to take account of the cognitive profile of adolescents.

However, this systematic review is not without its limitations. First, this review was conducted to understand the usability of adolescent health information websites only. With recent advances in technology, several digital interventions for adolescents have been designed with other media in mind. This includes platforms such as mobile apps (eg, [63,64], social media [3], and video games [72-74]). It is possible that the usability of these platforms is different than the usability of health information websites and should be systematically reviewed in further studies. Usability features of websites that were designed to change adolescents’ health behavior were also excluded from this review. These websites are likely to have unique usability features that were not typically present in health information websites (eg, tools to monitor health behavior, feedback on progress, tailored content, messages of encouragement, and motivation). It will be worthwhile to review the usability of websites that focus on changing health behavior in future studies. Second, this review did not evaluate the content of health information websites. In addition to good usability, it is also important that the content is age-appropriate, accurate, and based on quality evidence. Reviewing the content of the health information websites was beyond the scope of this review. Third, adolescents are a heterogeneous group. However, the differences between the younger and older adolescent group, and gender was only partially addressed in this review. It was not possible to evaluate the findings based on other demographics such as socioeconomic background and ethnicity because of limited information collected in the original studies, even though this could have influenced how web-based health information is accessed and interpreted [17]. In addition, it should be noted that studies were included even if some participants were aged over 18 years or under 13 years. It is possible that by including participants beyond the age range specified, we cannot confidently generalize our findings to the age group of 13 to 17 years. However, the age range of adolescents can vary considerably across studies, with participants aged between 12 and 24 years often considered to fall in the broad category of adolescence [17]. It was felt that by excluding studies strictly based on the age range of 13 to 17 years, studies may have been excluded that were still relevant to our central focus. Finally, this review was primarily based on qualitative studies for which a narrative synthesis was considered the most suitable. However, it should be noted that a qualitative review is subject to greater analysis bias than a quantitative review.

Conclusions

The broad insight offered by this systematic review is that adolescents have specific preferences related to the usability of health information websites. Adolescents generally prefer interactive content such as games and quizzes, as well as images, graphics, videos, and animations. These features are generally preferred if clearly presented, help separate large pages of text and do not clutter the website. Social networking features are also favored by adolescents, such as discussion boards and chat rooms. However, if health information websites are not able to provide these features, it is recommended that real stories and testimonials from other adolescents with the relevant health issue are made available on the website. Website design should take account of the preferences, skills, and neurodevelopmental profile of adolescents. If a clinical group with a specific cognitive profile is targeted, then this should also be accommodated. The findings of this review could help inform the development of more successful health information websites for adolescents.

Conflicts of Interest

GKR has no disclosures. LM has no disclosures. DWL's disclosures are as follows: DWL has participated in speaker bureau for Bayer, Merck, Almirall, Excemed, TEVA, Roche, Novartis, Biogen, Sanofi; has had consultancy from Novartis, Bayer, Merck, Biogen, TEVA, Sanofi; has had research grants from Bayer, Merck, Novartis, Biogen. All are paid into DWL's institution.

Multimedia Appendix 1

Quality assessment of included studies (n=25).

[DOCX File, 14KB - jmir_v21i4e11584_app1.docx]

Multimedia Appendix 2

Website characteristics and patient feedback about website usability (n=25).

[DOCX File, 32KB - jmir_v21i4e11584_app2.docx]

References

1. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478. [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
2. Stevens MC. The contributions of resting state and task-based functional connectivity studies to our understanding of adolescent brain network maturation. Neurosci Biobehav Rev 2016 Dec;70:13-32. [doi: 10.1016/j.neubiorev.2016.07.027] [Medline: 27502750]
3. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, Health and Technology: A National Survey. Media Commun 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
4. Borzekowski DL. Adolescents' use of the Internet: a controversial, coming-of-age resource. Adolesc Med Clin 2006 Feb;17(1):205-216. [doi: 10.1016/j.admecli.2005.10.006] [Medline: 16473301]
5. Franck LS, Noble G. Here's an idea: ask the users! Young people's views on navigation, design and content of a health information website. J Child Health Care 2007 Dec;11(4):287-297. [doi: 10.1177/1367493507083941] [Medline: 18039731]
6. Freeman JL, Caldwell PH, Bennett PA, Scott KM. How adolescents search for and appraise online health information: A systematic review. J Pediatr 2018 Apr;195:244-55.e1. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
7. Gray NJ, Klein JD, Noyce PR, 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]
8. Bröder J, Okan O, Bauer U, Bruland D, Schlupp S, Bollweg TM, et al. Health literacy in childhood and youth: a systematic review of definitions and models. BMC Public Health 2017 Dec 26;17(1):361 [FREE Full text] [doi: 10.1186/s12889-017-4267-y] [Medline: 28441934]
9. Jiménez-Pernett J, García-Gutiérrez JF, Bermúdez-Tamayo C, Labry-Lima AO. [Webliography of health for adolescents and young people]. Rev Calid Asist 2010;25(1):4-11. [doi: 10.1016/j.cali.2009.07.004] [Medline: 19837624]
10. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the internet: an observational study. J Med Internet Res 2003 Oct 17;5(4):e25 [FREE Full text] [doi: 10.2196/jmir.5.4.e25] [Medline: 14713653]
11. Alnemer KA, Alhuzaim WM, Alnemer AA, Alharbi BB, Bawazir AS, Barayyan OR, et al. Are health-related tweets evidence based? Review and analysis of health-related tweets on Twitter. J Med Internet Res 2015 Oct 29;17(10):e246 [FREE Full text] [doi: 10.2196/jmir.4898] [Medline: 26515535]
12. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
13. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? Ann Behav Med 2004 Feb;27(1):3-12. [doi: 10.1207/s15324796abm2701_2] [Medline: 14979858]
14. Bano M, Zowghi D. A systematic review on the relationship between user involvement and system success. Inf Softw Technol 2015 Feb;58(9):148-169 [FREE Full text] [doi: 10.1016/j.infsof.2014.06.011]
15. Dijkstra AA, Voorn P, Berendse HW, Groenewegen HJ, Rozemuller AJ, van de Berg WD, Netherlands Brain Bank. Stage-dependent nigral neuronal loss in incidental Lewy body and Parkinson's disease. Mov Disord 2014 Sep;29(10):1244-1251. [doi: 10.1002/mds.25952] [Medline: 24996051]
16. Rasulov AS, Evstigneeva ZG, Kretovich WL. [Irreversible inactivation of Chlorella glutamine synthetase by urea]. Biokhimiia 1978 May;43(5):912-918. [Medline: 26431]
17. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 3;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]

18. 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]
19. Sillence E, Briggs P, Harris P, Fishwick L. A framework for understanding trust factors in web-based health advice. Int J Hum Comput Stud 2006 Aug;64(8):697-713. [doi: 10.1016/j.ijhcs.2006.02.007]
20. Hannes K, Lockwood C, Pearson A. A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. Qual Health Res 2010 Dec;20(12):1736-1743. [doi: 10.1177/1049732310378656] [Medline: 20671302]
21. Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan M, et al. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. Soc Sci Med 2003 Feb;56(4):671-684. [doi: 10.1016/S0277-9536(02)00064-3] [Medline: 12560003]
22. Reen G, Silber E, Langdon D. Multiple sclerosis patients' understanding and preferences for risks and benefits of disease-modifying drugs: a systematic review. J Neurol Sci 2017 Apr 15;375:107-122. [doi: 10.1016/j.jns.2016.12.038] [Medline: 28320112]
23. Ammerlaan JJ, Scholtus LW, Drossaert CH, van Os-Medendorp H, Prakken B, Kruize AA, et al. Feasibility of a website and a hospital-based online portal for young adults with juvenile idiopathic arthritis: views and experiences of patients. JMIR Res Protoc 2015;4(3):e102 [FREE Full text] [doi: 10.2196/resprot.3952] [Medline: 26276373]
24. Baulch J, Chester A, Brennan L. Adolescent and parent content preferences and predictors of intention to use an online healthy weight website for adolescents. E-Journal Appl Psychol 2010 May 25;6(1):19-27. [doi: 10.7790/ejap.v6i1.173]
25. Breakey VR, Warias AV, Ignas DM, White M, Blanchette VS, Stinson JN. The value of usability testing for Internet-based adolescent self-management interventions:. BMC Med Inform Decis Mak 2013 Oct 4;13:113 [FREE Full text] [doi: 10.1186/1472-6947-13-113] [Medline: 24094082]
26. Breakey VR, Ignas DM, Warias AV, White M, Blanchette VS, Stinson JN. A pilot randomized control trial to evaluate the feasibility of an Internet-based self-management and transitional care program for youth with haemophilia. Haemophilia 2014 Nov;20(6):784-793. [doi: 10.1111/hae.12488] [Medline: 25311370]
27. Coyne I, Prizeman G, Sheehan A, Malone H, While AE. An e-health intervention to support the transition of young people with long-term illnesses to adult healthcare services: design and early use. Patient Educ Couns 2016 Dec;99(9):1496-1504. [doi: 10.1016/j.pec.2016.06.005] [Medline: 27372524]
28. Cullen KW, Thompson D, Boushey C, Konzelmann K, Chen T. Evaluation of a web-based program promoting healthy eating and physical activity for adolescents: teen choice: food and fitness. Health Educ Res 2013 Aug;28(4):704-714 [FREE Full text] [doi: 10.1093/her/cyt059] [Medline: 23748162]
29. Danielson C, McCauley J, Gros K, Jones A, Barr S, Borkman A, et al. SiHLEWeb.com: development and usability testing of an evidence-based HIV prevention website for female African-American adolescents. Health Informatics J 2016 Dec;22(2):194-208 [FREE Full text] [doi: 10.1177/1460458214544048] [Medline: 25167865]
30. DeBar L, Dickerson J, Clarke G, Stevens V, Ritenbaugh C, Aickin M. Using a website to build community and enhance outcomes in a group, multi-component intervention promoting healthy diet and exercise in adolescents. J Pediatr Psychol 2009 Jun;34(5):539-550 [FREE Full text] [doi: 10.1093/jpepsy/jsn126] [Medline: 19091807]
31. Donovan E, Wood M, Frayjo K, Black R, Surette D. A randomized, controlled trial to test the efficacy of an online, parent-based intervention for reducing the risks associated with college-student alcohol use. Addict Behav 2012 Jan;37(1):25-35 [FREE Full text] [doi: 10.1016/j.addbeh.2011.09.007] [Medline: 21963316]
32. Ercan S. Evaluation of a mental health website for teenagers. Psychiatr Bull 2006 May 1;30(5):175-178. [doi: 10.1192/pb.30.5.175]
33. Hanberger L, Ludvigsson J, Nordfeldt S. Use of a web 2.0 portal to improve education and communication in young patients with families: randomized controlled trial. J Med Internet Res 2013;15(8):e175 [FREE Full text] [doi: 10.2196/jmir.2425] [Medline: 23973555]
34. Korus M, Cruchley E, Stinson JN, Gold A, Anthony SJ. Usability testing of the Internet program:. Pediatr Transplant 2015 Feb;19(1):107-117. [doi: 10.1111/petr.12396] [Medline: 25495484]
35. Long AC, Palermo TM. Brief report: web-based management of adolescent chronic pain: development and usability testing of an online family cognitive behavioral therapy program. J Pediatr Psychol 2009 Jun;34(5):511-516 [FREE Full text] [doi: 10.1093/jpepsy/jsn082] [Medline: 18669578]
36. McCarthy O, Carswell K, Murray E, Free C, Stevenson F, Bailey JV. What young people want from a sexual health website: design and development of Sexunzipped. J Med Internet Res 2012;14(5):e127 [FREE Full text] [doi: 10.2196/jmir.2116] [Medline: 23060424]
37. Michaud P, Colom P. Implementation and evaluation of an internet health site for adolescents in Switzerland. J Adolesc Health 2003 Oct;33(4):287-290. [Medline: 14519571]
38. Nicholas D, Fellner K, Frank M, Small M, Hetherington R, Slater R, et al. Evaluation of an online education and support intervention for adolescents with diabetes. Soc Work Health Care 2012;51(9):815-827. [doi: 10.1080/00981389.2012.699507] [Medline: 23078013]

39. Nordfeldt S, Hanberger L, Berterö C. Patient and parent views on a Web 2.0 Diabetes Portal--the management tool, the generator, and the gatekeeper: qualitative study. J Med Internet Res 2010;12(2):e17 [FREE Full text] [doi: 10.2196/jmir.1267] [Medline: 20511179]
40. Radovic A, DeMand A, Gmelin T, Stein B, Miller E. SOVA: design of a stakeholder informed social media website for depressed adolescents and their parents. J Technol Hum Serv 2018;35(3):169-182 [FREE Full text] [doi: 10.1080/15228835.2017.1347552] [Medline: 29743822]
41. Radovic A, Gmelin T, Hua J, Long C, Stein BD, Miller E. Supporting Our Valued Adolescents (SOVA), a social media website for adolescents with depression and/or anxiety: technological feasibility, usability, and acceptability study. JMIR Ment Health 2018 Feb 26;5(1):e17 [FREE Full text] [doi: 10.2196/mental.9441] [Medline: 29483067]
42. Starling R, Nodulman J, Kong A, Wheeler C, Buller D, Woodall W. Usability testing of an HPV information website for parents and adolescents. Online J Commun Media Technol 2015 Oct;5(4):184-203 [FREE Full text] [Medline: 26594313]
43. Stinson J, Gupta A, Dupuis F, Dick B, Laverdière C, LeMay S, et al. Usability testing of an online self-management program for adolescents with cancer. J Pediatr Oncol Nurs 2015;32(2):70-82. [doi: 10.1177/1043454214543021] [Medline: 25037173]
44. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
45. Stinson JN, McGrath PJ, Hodnett ED, Feldman BM, Duffy CM, Huber AM, et al. An internet-based self-management program with telephone support for adolescents with arthritis: a pilot randomized controlled trial. J Rheumatol 2010 Sep;37(9):1944-1952. [doi: 10.3899/jrheum.091327] [Medline: 20595280]
46. Wozney L, Baxter P, Newton A. Usability evaluation with mental health professionals and young people to develop an Internet-based cognitive-behaviour therapy program for adolescents with anxiety disorders. BMC Pediatr 2015;15:213. [doi: 10.1186/s12887-015-0534-1] [Medline: 26675420]
47. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. The internet: a window on adolescent health literacy. J Adolesc Health 2005 Sep;37(3):243. [doi: 10.1016/j.jadohealth.2004.08.023] [Medline: 16109345]
48. Ståhl T. How ICT savvy are digital natives actually? Nord J Digit Lit 2017;12(3):89-108. [doi: 10.18261/issn.1891-943x-2017-03-04]
49. Uncapher MR, Lin L, Rosen LD, Kirkorian HL, Baron NS, Bailey K, et al. Media multitasking and cognitive, psychological, neural, and learning differences. Pediatrics 2017 Nov 1;140(Supplement 2):S62-S66. [doi: 10.1542/peds.2016-1758D]
50. Giedd JN. The digital revolution and adolescent brain evolution. J Adolesc Health 2012 Aug;51(2):101-105 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.002] [Medline: 22824439]
51. Mills KL, Goddings A, Herting MM, Meuwese R, Blakemore S, Crone EA, et al. Structural brain development between childhood and adulthood: convergence across four longitudinal samples. Neuroimage 2016 Nov 1;141:273-281 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.07.044] [Medline: 27453157]
52. Göllner LM, Ballhausen N, Kliegel M, Forstmeier S. Delay of gratification, delay discounting and their associations with age, episodic future thinking, and future time perspective. Front Psychol 2017;8:2304 [FREE Full text] [doi: 10.3389/fpsyg.2017.02304] [Medline: 29422875]
53. Mueller SC, Cromheeke S, Siugzdaite R, Nicolas BC. Evidence for the triadic model of adolescent brain development: cognitive load and task-relevance of emotion differentially affect adolescents and adults. Dev Cogn Neurosci 2017 Dec;26:91-100 [FREE Full text] [doi: 10.1016/j.dcn.2017.06.004] [Medline: 28688343]
54. Padilla-Walker LM, Carlo G, Memmott-Elison MK. Longitudinal change in adolescents' prosocial behavior toward strangers, friends, and family. J Res Adolesc 2017 Nov 16:698-710. [doi: 10.1111/jora.12362] [Medline: 29144027]
55. Sebastian C, Viding E, Williams K, Blakemore S. Social brain development and the affective consequences of ostracism in adolescence. Brain Cogn 2010 Feb;72(1):134-145. [doi: 10.1016/j.bandc.2009.06.008] [Medline: 19628323]
56. Crone E, Konijn E. Media use and brain development during adolescence. Nat Commun 2018 Dec 21;9(1):588 [FREE Full text] [doi: 10.1038/s41467-018-03126-x] [Medline: 29467362]
57. Kaye S, White M, Lewis I. The use of neurocognitive methods in assessing health communication messages: a systematic review. J Health Psychol 2017 Dec;22(12):1534-1551. [doi: 10.1177/1359105316630138] [Medline: 26908587]
58. Happé F, Frith U. Annual research review: towards a developmental neuroscience of atypical social cognition. J Child Psychol Psychiatry 2014 Jun;55(6):553-557. [doi: 10.1111/jcpp.12162] [Medline: 24963529]
59. DuBois D, Lymer E, Gibson B, Desarkar P, Nalder E. Assessing sensory processing dysfunction in adults and adolescents with autism spectrum disorder: a scoping review. Brain Sci 2017 Aug 19;7(8):108 [FREE Full text] [doi: 10.3390/brainsci7080108] [Medline: 28825635]
60. Ekmekci O. Pediatric multiple sclerosis and cognition: a review of clinical, neuropsychologic, and neuroradiologic features. Behav Neurol 2017;2017:1463570 [FREE Full text] [doi: 10.1155/2017/1463570] [Medline: 29434433]
61. Imhof M, Schmälzle R, Renner B, Schupp H. How real-life health messages engage our brains: shared processing of effective anti-alcohol videos. Soc Cogn Affect Neurosci 2017 Dec 1;12(7):1188-1196 [FREE Full text] [doi: 10.1093/scan/nsx044] [Medline: 28402568]

62. Raschle NM, Fehlbaum LV, Menks WM, Euler F, Sterzer P, Stadler C. Investigating the neural correlates of emotion–cognition interaction using an affective stroop task. Front Psychol 2017;8:1489 [FREE Full text] [doi: 10.3389/fpsyg.2017.01489] [Medline: 28919871]
63. Kenny R, Dooley B, Fitzgerald A. Feasibility of. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
64. Boulos MNK, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
65. Reichow B, Gelbar NW, Mouradjian K, Shefcyk A, Smith IC. Characteristics of international websites with information on developmental disabilities. Res Dev Disabil 2014 Oct;35(10):2293-2298. [doi: 10.1016/j.ridd.2014.05.028] [Medline: 24952371]
66. Grant N, Rodger S, Hoffmann T. Evaluation of autism-related health information on the web. J Appl Res Intellect Disabil 2015 Jul;28(4):276-282 [FREE Full text] [doi: 10.1111/jar.12127] [Medline: 25224690]
67. Manting EH, Driessen AJ. Escherichia coli translocase: the unravelling of a molecular machine. Mol Microbiol 2000 Jul;37(2):226-238 [FREE Full text] [doi: 10.1046/j.1365-2958.2000.01980.x] [Medline: 10931320]
68. Berk L, Berk M, Dodd S, Kelly C, Cvetkovski S, Jorm AF. Evaluation of the acceptability and usefulness of an information website for caregivers of people with bipolar disorder. BMC Med 2013 Jul 11;11:162 [FREE Full text] [doi: 10.1186/1741-7015-11-162] [Medline: 23844755]
69. Kushalnagar P, Naturale J, Paludneviciene R, Smith S, Werfel E, Doolittle R, et al. Health websites: accessibility and usability for American sign language users. Health Commun 2015;30(8):830-837 [FREE Full text] [doi: 10.1080/10410236.2013.853226] [Medline: 24901350]
70. Ryan G, Haroon M, Melvin G. Evaluation of an educational website for parents of children with ADHD. Int J Med Inform 2015 Nov;84(11):974-981. [doi: 10.1016/j.ijmedinf.2015.07.008] [Medline: 26265480]
71. Scanlan F, Jorm A, Reavley N, Meyer D, Bhar S. Treatment choices for depression: young people's response to a traditional e-health versus a Health 2.0 website. Digit Health 2017;3:2055207617690260 [FREE Full text] [doi: 10.1177/2055207617690260] [Medline: 29942581]
72. Hale A, Young V, Grand A, McNulty C. Can gaming increase antibiotic awareness in children? A mixed-methods approach. JMIR Serious Games 2017 Mar 24;5(1):e5 [FREE Full text] [doi: 10.2196/games.6420] [Medline: 28341618]
73. Farrell D, Kostkova P, Weinberg J, Lazareck L, Weerasinghe D, Lecky DM, et al. Computer games to teach hygiene: an evaluation of the e-Bug junior game. J Antimicrob Chemother 2011 Jun;66(Suppl 5):v39-v44 [FREE Full text] [doi: 10.1093/jac/dkr122] [Medline: 21680586]
74. Sparapani VD, Fels S, Nascimento LC. The value of children's voices for a video game development in the context of type 1 diabetes: focus group study. JMIR Diabetes 2017 Jul 19;2(2):e17 [FREE Full text] [doi: 10.2196/diabetes.7652] [Medline: 30291061]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CASP: Critical Appraisal Skills Programme
ERIC: Education Resources Information Center

Edited by G Eysenbach; submitted 13.07.18; peer-reviewed by E Park, P Kreniske; comments to author 30.08.18; revised version received 12.12.18; accepted 11.02.19; published 23.04.19.

Please cite as:
Reen GK, Muirhead L, Langdon DW
Usability of Health Information Websites Designed for Adolescents: Systematic Review, Neurodevelopmental Model, and Design Brief
J Med Internet Res 2019;21(4):e11584
URL: http://www.jmir.org/2019/4/e11584/
doi:10.2196/11584
PMID:31012856

Original Paper

Using the Technology Acceptance Model to Explore User Experience, Intent to Use, and Use Behavior of a Patient Portal Among Older Adults With Multiple Chronic Conditions: Descriptive Qualitative Study

Jennifer Dickman Portz[1,2], MSW, PhD; Elizabeth A Bayliss[2,3], MD, MPH; Sheana Bull[4], PhD; Rebecca S Boxer[2], MD; David B Bekelman[1,5], MD; Kathy Gleason[2], PhD; Sara Czaja[6,7], PhD

[1]Division of General Internal Medicine, School of Medicine, University of Colorado, Aurora, CO, United States
[2]Institute for Health Research, Kaiser Permanente Colorado, Aurora, CO, United States
[3]Department of Family Medicine, School of Medicine, University of Colorado, Aurora, CO, United States
[4]mHealth Impact Lab, Colorado School of Public Health, University of Colorado, Aurora, CO, United States
[5]Department of Medicine, Eastern Colorado Health Care System, Department of Veterans Affairs, Denver, CO, United States
[6]Division of Geriatrics, Weill Cornell Medicine, New York, NY, United States
[7]Center for Research and Education on Aging and Technology Enhancement, University of Miami, Miami, FL, United States

Corresponding Author:
Jennifer Dickman Portz, MSW, PhD
Division of General Internal Medicine
School of Medicine
University of Colorado
Mailstop B119
13001 East 17th Place
Aurora, CO, 80045
United States
Phone: 1 303 724 4438
Email: jennifer.portz@colostate.edu

Abstract

Background: Patient portals offer modern digital tools for older adults with multiple chronic conditions (MCC) to engage in their health management. However, there are barriers to portal adoption among older adults. Understanding portal user interface and user experience (UI and UX) preferences of older adults with MCC may improve the accessibility, acceptability, and adoption of patient portals.

Objective: The aim of this study was to use the Technology Acceptance Model (TAM) as a framework for qualitatively describing the UI and UX, intent to use, and use behaviors among older patients with MCC.

Methods: We carried out a qualitative descriptive study of Kaiser Permanente Colorado’s established patient portal, *My Health Manager*. Older patients (N=24; mean 78.41 (SD 5.4) years) with MCC participated in focus groups. Stratified random sampling was used to maximize age and experience with the portal among participants. The semistructured focus groups used a combination of discussion and think-aloud strategies. A total of 2 coders led the theoretically driven analysis based on the TAM to determine themes related to use behavior, portal usefulness and ease of use, and intent to use.

Results: Portal users commonly used email, pharmacy, and lab results sections of the portal. Although, generally, the portal was seen to be easy to use, simple, and quick, challenges related to log-ins, UI design (color and font), and specific features were identified. Such challenges inhibited participants’ intent to use the portal entirely or specific features. Participants indicated that the portal improved patient-provider communication, saved time and money, and provided relevant health information. Participants intended to use features that were beneficial to their health management and easy to use.

Conclusions: Older adults are interested in using patient portals and are already taking advantage of the features available to them. We have the opportunity to better engage older adults in portal use but need to pay close attention to key considerations promoting usefulness and ease of use.

(J Med Internet Res 2019;21(4):e11604) doi:10.2196/11604

KEYWORDS

multiple chronic conditions; personal health record; patient portals; aging; health information technology

Introduction

Background

Patient portals, also referred to as tethered personal health records, are secure websites for personal health information and patient resources directly linked to a provider's electronic health record [1,2]. Patient portals offer modern digital tools for older adults to engage in health management and with their health care system [3]. In the United States, patient portals vary greatly by provider and health care system, but often provide access to personal health information, email messaging with providers, appointment schedulers, and prescription management [1]. Patient portals are designed to help patients better manage their health with the intention to improve health outcomes, health care communication, and reduce costs [4,5]. With access to lab results and health indicators such as weight, blood pressure, and cholesterol, patients can promote early intervention when they encounter a deviation or problem or monitor improvements if following a new medication, exercise, or diet regimen [6]. Patients can also access information from various providers and necessary medical histories during emergencies to improve care transition coordination [7]. Patient portals can be convenient for medication refills, scheduling appointments, and allowing patients to communicate asynchronously with providers [6,8]. Owing to consumer demand and US government incentives for health information technology expansion, the adoption and use of patient portals is on the rise [9].

Patient portals are a promising but understudied clinical tool particularly in aging populations [10]. As older adults are more likely to have multiple chronic conditions (MCC) and higher health care utilization [11,12], they are likely to benefit from patient portal use to manage their conditions and health care services [13,14]. Although older adults are the fastest growing users of the internet [15,16], a lag in patient portal adoption remains, particularly among the oldest, less affluent, and lower educated older adults [17,18]. Although older adults show interest in patient portals [19-21], adoption barriers and low utilization have been identified [22]. Older adults pinpoint technology discomfort, privacy and security concerns, and lack of relative advantages as primary reasons for not using patient portals [22,23].

Understanding older adult opinions about portal user interface (UI) and user experience (UX) may lead to improvements in the accessibility, acceptability, and adoption of patient portals among older adults with MCC. UI typically focuses on the visual *look* of the design, including elements related to color, font, and images. UX targets the overall experience related to usability, usefulness, function, credibility, and satisfaction with the technology [24]. There are UI design recommendations for older users [25]; however, little is known about portal use and UX among older adults with MCC [26]. For example, portal email communication, lab results access, and electronic refill capabilities are important features for portal users [18,27], yet, it is unknown if these tools are commonly used or valued specifically by older adults with MCC. As older adults with MCC have much to gain from using portals, research is needed to better understand use behavior, perceived benefits, and strategies for increasing portal use among this specific population [26]. Therefore, the purpose of this study was to qualitatively explore perspectives from older adults with MCC regarding Kaiser Permanente Colorado's (KPCO) established patient portal, *My Health Manager*. Framed by the Technology Acceptance Model (TAM), we qualitatively described the UI/UX, intent to use, and use behaviors among older Kaiser patients with MCC. Our specific research questions included the following: (1) How do participants use the portal?; (2) Why do participants use (or not use) the portal as they describe?; (3) How is the portal useful and usable?; and (4) How do these opinions and experiences influence participants' intent to use the portal?

Technology and Acceptance Model

The TAM is an information technology framework for understanding users' adoption and use of emerging technologies particularly in the workplace environment and has been tested in older populations [28,29]. The theory posits that a person's *intent to use* (acceptance of technology) and *usage behavior* (actual use) of a technology is predicated by the person's perceptions of the specific technology's *usefulness* (benefit from using the technology) and *ease of use*. Simply, users are more likely to adopt a new technology with high-quality UX design (ie, usable, useful, desirable, and credible). The TAM also suggests that perceptions of usefulness and ease of use are mediated by *external variables* including individual differences, system characteristics, social influences, and facilitating conditions.

Kaiser Permanente Patient Portal: My Health Manager

My Health Manager (Table 1) provides personal health information related to patient diagnosis, prescriptions, laboratory results, and vaccination records. To improve provider-patient communication, *My Health Manager* offers features for patients to email providers and schedule appointments. Health management features that are designed to foster healthy eating and exercise habits incorporate personalized assessments and health self-management tools.

Table 1. Patient portal (My Health Manager) features summary.

Feature	Function
Appointment Center	Patients can schedule or cancel appointments
My Medical Record	Patients can view test results, immunization records, medical problem list, and care plans
Pharmacy Center	Patients can manage prescriptions and order medications
Health Guides and Health Management Tools	Access to health resources and self-management tools for diet, exercise, smoking cessation, and disease specific care
Message Center	Patients can email with their provider
Recently added features	e-visit and provider *chat* functions for nonemergent questions and visits

Methods

This is an exploratory, descriptive qualitative study based on data collected from a series of focus groups [30,31]. This method is primarily used to better understand the needs and desired outcomes from a particular group of people [32]. As such, this study employed an exploratory, descriptive approach to describe UI and UX preferences, intent to use, and use behaviors of KPCO's *My Health Manager* of older patients with MCC . All procedures were approved by the KPCO Institutional Review Board.

Sample and Recruitment

We identified KPCO patients meeting the following inclusion criteria: aged ≥65 years, KPCO member for ≥1 year, presence of MCC (Charlson Comorbidity Index >2), and connected to 1 of 3 clinics in the Denver metro area with large geriatric patient populations. Non-English–speaking patients, individuals residing in skilled nursing facilities, and patients with a diagnosis of dementia were excluded. We then randomly selected potential participants stratified by age group (65 to 75 years; 76 to 85 years; and 86+ years) and portal user status (nonusers and users) to ensure participation from older participants and maximize the variability of experience with the portal. *Nonusers* were patients not registered for the portal, and *users* were those registered for the portal and logged into the portal within the last 6 months. Recruitment letters were mailed to 225 potential participants summarizing the study and providing an opt-out phone number to call if disinterested. A total of 210 potential participants (n=90 users and n=120 nonusers), who did not initially opt-out, were contacted via phone and invited to participate in focus groups. Recruitment resulted in an 18% acceptance rate (n=19 users and n=18 nonusers). Of the 37 patients that were scheduled to attend focus groups, 24 patients (n=15 users; n=9 nonusers) participated. We contacted the 13 participants who did not show up for their scheduled focus group to reschedule: 2 patients were unable to reschedule and 11 were lost to follow-up.

Focus Groups and Question Guide

We conducted 6 focus groups (3 nonuser groups and 3 user groups) lasting approximately 90 min at the KPCO facility most convenient for the participant. Around 3 to 6 patients participated in each group. Focus groups were semistructured in format, allowing for probing and extended discussion on topics of interest to the participants. Before group discussion, participants were asked to complete a demographics and technology utilization survey. The survey collected information regarding income, education, cell phone, email, internet, digital communication, and social media use. Other demographic variables including race/ethnicity, age, and days since last portal log-in were captured from KPCO's electronic medical record during sampling procedures. Participants were then asked questions related to *My Health Manager*. In addition to a traditional question answer session, patients were asked to *think aloud* [33] as a group, whereas the interviewer navigated a mock *My Health Manager* portal (Textboxes 1 and 2). This method has previously been used to assess health literacy and numeracy of patient portals among patients over 65 years [34]. Focus groups were audio-recorded for accuracy in data.

Data Analysis

A theoretically driven approach [35] based on the TAM was used for analysis to capture participants' opinions and experiences. The analysis was completed by 2 female doctoral-level researchers: JDP, a social work assistant professor with prior qualitative experience, and KG, a research assistant new to qualitative approaches. Audio files were first professionally transcribed verbatim. The unit of analysis, defined as a completion of 1 thought, was determined by the analytic team. Units of analysis ranged from a brief 3-word sentence to a paragraph of 6 sentences and were entered into Microsoft Excel to ensure that the units were consistent across coders. For initial coding, coders used a TAM-based theoretically driven code book developed *a priori* to code the units. The codebook included a list of TAM codes (eg, TAM constructs related to user intent, usage behavior, usefulness, and ease of use), code meaning, and criteria for using each code to capture participants' perceptions of their intent to use and UX of *My Health Manager*. To prevent coders from forcing the units to fit with the TAM framework, coders used a *no code* option for responses that did not meet the code meaning and criteria for the *a priori* codes.

Textbox 1. Overview of focus group questions and portal features for users.

Preliminary questions:

- Why did you enroll in My Health Manager?
- What features do you use most?

Think Aloud Questions:

- As we navigate this feature, what do you think about it?
- Why do you use it?
- How would you improve this feature?
- What helps you use this feature?
- What outcome do you want to achieve by using the feature?

Textbox 2. Overview of focus group questions and portal features for nonusers.

Preliminary questions:

- Are you interested in using My Health Manager?
- Are there reasons why you do not use My Health Manager?

Think Aloud Questions:

- As we navigate the portal, are there features you might like to use?
- What do you like about this feature? Or What do you dislike about this feature?
- Why would you want to use these features? Or Why would you prefer to NOT use these features?
- What supports would you need to use this feature?
- How do you currently accomplish this task (related to feature)?

However, to ensure that participants' thoughts were fully captured, coders also used a combination of *open* and *in vivo* coding (the use of participants' own words as a code) to add inductive codes to the code book as needed. Inter-rater reliability was calculated for all transcripts (K=0.98), reflecting adequate consistency in coding across coders. Patterned coding was then employed on initial codes to identify (1) patterns in responses between users and nonusers and (b) patterns in responses between ease of use, perceived usefulness, intent to use, and UX. These patterns were then used to form themes related to each research question and develop overall findings. The analytics team met regularly through the analytic process to discuss codes and correct any disagreements in coding and thematic findings.

Results

Participants

Participants (N=24) were of a mean age of 78 years and were primarily white women (Table 2). Patient portal users had logged in to *My Health Manager* on an average 17.1 (SD 28.3) days before recruitment. All but one participant used a cell phone regularly, primarily a mobile phone. The majority of participants, regardless of user status, used email and looked up information on the Web. Approximately half of the participants used social media, played video games, and used video chat, whereas instant messaging was less popular.

Technology Acceptance Model Description for My Health Manager

On the basis of the TAM, Figure 1 illustrates the findings from the focus groups regarding UI and UX, intent to use, and use behavior for My Health Manager.

Use Behavior

Portal users described their use of various *My Health Manager* features (listed in Table 1). The email Message Center was the most popular feature used by *My Health Manager* users. Of the participants, 1 stated, "Yeah, I email my doctors a lot!" whereas another stated, "And I like text chatting with the sending an e-mail to my doctors [feature]; just to ask a question". The Pharmacy Center was also commonly used to refill medications. For example, a participant said, "I use the pharmacy part every time. I hardly ever call in the pharmacy any more". Viewing lab results in the My Medical Record page was frequently used. As 1 participant noted, "I especially like looking up the results of my test and finding out what those tests are for and if there [is] anything I need to be concerned about."

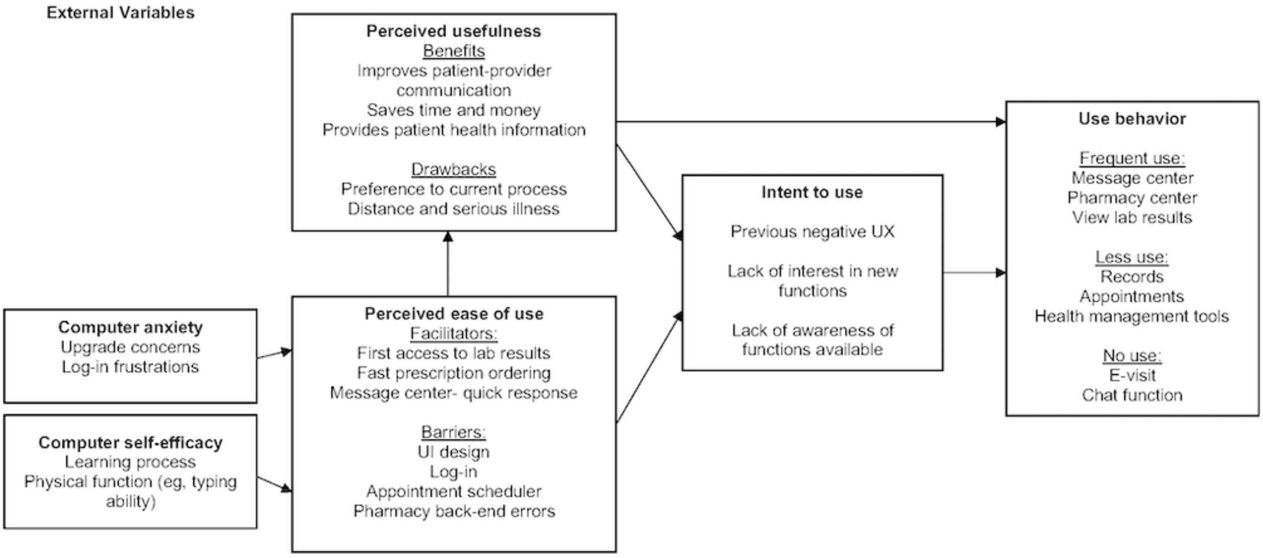
Participants did not commonly use other My Medical Record features including viewing diagnosis list, care plans, or immunizations records. In terms of using the Appointment Center to schedule visits, 1 participant explained, "I really like using the website. I have made appointments and been shocked when I got them the next day on the computer. So sometimes I try to check there first, and then I'll call if I feel like I need to come in and I can't". However, some participants unsuccessfully tried to use the Appointment Center and the majority of participants called the Kaiser phone line to schedule appointments. Only 1 participant stated they used the health guides and health management tools: "I think that's a pretty neat feature. And then I've looked up things for my friends when they have questions." None of the participants had used the newly added *My Health Manager* features including the e-visit or provider chat.

Table 2. Participant characteristics.

Characteristics	Users (N=15)	Nonusers (N=9)	Total (N=24)
Age (years), mean (SD)	76.4 (4.9)	82.7 (3.4)	78.41 (5.4)
Female, n (%)	12 (80)	5 (56)	17 (71)
White, n (%)	12 (80)	7 (78)	19 (79)
Hispanic, n (%)	1 (7)	2 (11)	3 (13)
Days since *My Health Manager* log in, mean (SD)	17.1 (28)	—[a]	—
Education, n (%)			
High School Graduate	1 (7)	5 (56)	6 (25)
Some College Graduate	7 (47)	2 (22)	9 (36)
College Graduate	7 (47)	2 (22)	9 (36)
Income (US$), n (%)			
<$30,000	2 (13)	2 (22)	4 (17)
$30,000 to $49,999	7 (47)	6 (67)	13 (54)
$50,000 to $74,999	2 (13)	0 (0)	2 (8)
$75,000 and more	2 (13)	0 (0)	2 (8)
Choose not to answer	2 (13)	1 (11)	3 (13)
Own cell phone, n (%)			
Smartphone	12 (80)	5 (56)	17 (71)
A regular or basic phone	12 (80)	3 (33)	6 (25)
Does not have a cell phone	3 (20)	1 (11)	1 (4)
Technology utilization, n (%)			
Email	15 (100)	7 (78)	22 (92)
Look up information on the Web	15 (100)	6 (67)	21 (88)
Use social media	8 (53)	5 (56)	13 (54)
Play computer games	12 (80)	3 (33)	15 (63)
Video chat	7 (47)	4 (44)	11 (46)
Instant messaging	6 (40)	2 (22)	8 (33)

[a] Not applicable.

Figure 1. Technology Acceptance Model description for My Health Manager. UX: user experience; UI: user interface.



Perceived Ease of Use

Although participants stated that the website was “pretty easy to use,” both portal user and nonuser participants (as shown during the focus group) were relatively negative about the UI and UX of *My Health Manager* (Textbox 3). Nonusers quickly identified UI design problems related to font size and colors while viewing the mock portal. In addition to some design issues, portal users noted challenges with using the portal or follow-up from using the portal system specifically related to registering with the system, logging in, and scheduling appointments. Several challenges were related to the back end of the system. For example, it was easy for patients to order their prescriptions on the Web, but in some cases, there were problems when participants went to the clinic to pick up the prescription. Participants who tried to use the appointment center said that they could not figure out how to schedule a visit on the portal. Those that were able to schedule an appointment on the portal experienced back-end problems at the clinic when they arrived to check-in for the visit. Despite some challenges in using the portal, many users thought *My Health Manager* was easy to use, and in some cases, *My Health Manager* was easier to use than more traditional services.

Perceived ease of use was impacted by participants’ level of computer anxiety and computer self-efficacy.

Computer Anxiety

Both portal users and nonusers expressed issues of computer anxiety, defined as apprehension or even fear of using the technology, which led to specific difficulties in using *My Health Manager*. As 1 user stated, “Well, I have just a couple of general thoughts about the Kaiser website. One is that what we all hate is instant upgrades, right? You hear the word ‘upgrade’ and you shudder.” Users worried that with every update, at some point they would be unable to use *My Health Manager* and that each new upgraded version created new use challenges. For example, during an upgrade, the log in page was modified, making it difficult for users to find the username and password text box that had previously been easy to use. The participant explained, “So recently—I don’t know how recently—but Kaiser upgraded, and all of a sudden, I couldn’t figure out how to log in. Now how stupid is that? So eventually, you know, I scanned and the ‘oh, there’s my name in the little arrow,’ but I thought, ‘couldn't they have warned us?’ That was my first thought.”

Therefore, after an upgrade, users were apprehensive about using the new version and typically found the new version to be more difficult to use than the previous version.

Computer Self-Efficacy

Users and nonusers both indicated that they were pretty confident in their ability to “figure things out” on the portal and felt good when they learned how to use a new *My Health Manager* feature or technology in general. However, participants discussed that although confident in their ability to use the portal, the process of learning how to use the website was challenging. A user stated:

> *Me and computers have problems anyway. It's like, ugh. Probably just inexperience of using the computer. I mean, I use the things that they have here, but it's not anything like going onto the website and that kind of stuff. But it was probably just not knowing where I was going or what I was doing.*

Responses suggested that older participants believed they could use *My Health Manager* but acknowledged there would be a steep learning curve. Although learning to use the portal presented difficulties, some participants explained they simply did not have the ability to use specific features. One nonuser, who was interested in learning more about the provider email function, stated they could not use the new provider chat function because they could not type or respond quickly enough to instant messages with a provider. They explained, “You have to type in your chat...It’s fast...So email would be better.”

Perceived Usefulness

Participants described benefits and drawbacks regarding My Health Manager (Textbox 4).

Textbox 3. Perceived ease of use of My Health Manager.

Participant quotes:

- Barriers
- User interface design

  It’s hard for me to see.

  The big print I can read. But I have to strain to see the small prints.

  It’s like it’s not dark enough. Is that what other people say, too?
- Logging-in

  My frustration is, it does not matter what I try to use, it never works. I can do this, this, this, this, this just exactly the way I remember the way I’m supposed to do it, and then you get down and it does not work. Enter password, wrong. This or that, something is wrong.
- Scheduling appointments

  Nope, I couldn’t make it work...Because you don’t know what’s available. Does (the doctor) have a slot at 2:00 on such-and-such a day? They don’t tell you that... Step 7? Why is there seven steps just to make an appointment?
- Back-end errors

  I don’t want to use (the Appointment Center) because there’s confusion among the people here for instance. And I’ll tell you, you make (an appointment), for instance I have sun damage because I'm out in the sun a lot, so I made an appointment to have it checked. And so that was the appointment. When I got here, to see the doctor, the nurse says, “Well, you have to see a PA.” I said, “Well, I just made an appointment on the website. I made the appointment, I wrote it down.”

  Well, yeah. They went through a period of time, I think, when they were changing over which that’s really the only problem I’ve ever had with them. And they were significant because medications disappeared. I mean you order them, they were there. But other than that, it works very well. They straightened it out I think.
- Facilitators
- Simple, quick, and easy to use

  It’s nice to be able to see the results and that stuff because that’s easy. Or if the doctor sends you a message. It’s easy to pick it up.

  I have sent emails to my doc, especially when I don't want to come in and usually I get an answer within a very short time.

  The test results are pretty easy to get. I really like it. I like the fact that I can graph my test results as opposed to just seeing the numbers.

Textbox 4. Perceived usefulness of My Health Manager.

Participant quotes:

- Benefits of using *My Health Manager*
- Improves patient-provider communication

  I love the fact that I can communicate with the doctor or any of the other doctors. I get complete descriptions on blood work and what happens with that.

  You can as you are typing (via email feature), you can think and maybe, “No, that isn’t really what I want to say” instead of stumbling around. And you can do it more precise...Yeah, more organized. And then before you send it.

  For me, it was an easy way to get non-emergent information to the doctor. For me, that's the easiest part of it is I can send stuff and they’ll either answer me or give me a call, one of the two.
- Saves time and money

  And that makes sense, because all it is, is you don’t have to come in for that visit, which, if it’s difficult for you to get out, if the weather’s crappy and stuff like that.

> *When you initially send the e-mail to your doctor, sometimes they get back to you and they'll say, "We'll have a conference call. I've arranged a conference call to talk to you about it." And that really saves a lot of time.*
>
> *And it could be something just little or a prescription change or something to that effect that you really don't have to come in and see the doctor about. And it's more an efficient way of really the whole system working.*

- Provides patient health information

> *I am an advocator for people taking control of their own health care versus relying on – that's not to say I'm going to self-medicate or anything. But I believe in being well-informed about my healthcare and presenting options to my doctors and that sort of thing. So I like to be really informed about what's going on.*
>
> *But yeah, they don't have any trouble because it's nice that I can get messages from my doctors, telling me where I'm at. Or if I've had a blood test, I know that it's okay or if it's not*

- Drawbacks of using *My Health Manager*
- Preference for current process

> *That's hard. There's yes and then no, because I don't hardly call into the hospital or I know when my appointments are and when I come, they tell me to call in to get my medicine. I don't because I live so close, it's even a little walk for me to come and pick them up. So, the way I feel now, I can still do things like I'm doing now. I mean you have to walk four or five blocks to come down here, and then I take what I can do by myself. And so, then sometimes I meet people here that I know, and for me it's just like getting out for a little trip.*

- Distance and serious illness

> *I think if I had more in my body or that I had more problems that (using My Health Manager) would be good. But I am never sick. Of course, you never want something to happen. I don't go to my doctor real often either*
>
> *If I lived far, it would work very good for me.*

Benefits

Portal users expressed a clear benefit in using *My Health Manager*, in general and for very specific features. Generally speaking, users expressed that *My Health Manager* was useful in communicating with their provider, accessing health information, saving time and money, and addressing health concerns without a clinic visit.

Drawbacks

On the contrary, nonusers stated they preferred to accomplish health-related tasks using their current process and indicated the portal would be more helpful for particular people: those living at a distance and those with serious illness. Nonusers preferred to use the telephone or clinic pharmacy for prescriptions, seeing their provider in person when asking nonemergent questions and calling to make clinical appointments. However, it is important to note that participants described alternative benefits for their current methods. Although portal users thought using *My Health Manager* for prescription refills was useful in getting medications, nonusers liked going to the pharmacy in person because it got them out of the house and kept them active. Nonusers acknowledged that *My Health Manager* would be useful for patients *far away* from their providers and *sicker* patients. Nonusers understood why people would want to use the system. It simply did not seem useful to them personally.

Intent to Use

Participants' intention to continue using or start using *My Health Manager* was influenced by their perceived ease of use and perceived usefulness.

Previous Negative User Experience

Once a participant had a negative UI with *My Health Manager* or a specific feature, they had little interest in trying again. For example, a nonuser who tried to register for *My Health Manager* was so frustrated with their registration experience that he/she did not want to try again, and gave the following explanation:

> *I tried to get on [My Health Manager] several years ago and that was when they were sending the password by mail. I lost the password, and I forgot that you had to have (a password). I forgot all that. I again tried to get on it and didn't have a password so I thought, "Well, I'll just start over again." It wouldn't let me, so I said, "Well, the heck with it then.*

Also, as stated above, users expressed many challenges in using the Appointment Center; therefore, participants showed little intent to use the Appointment Center until the *glitches* were fixed.

Lack of Interest in New Functions

Nonusers and users alike were not interested in using the recently added features of e-visits and chat functions. Few participants saw the value in using these features, articulating comments such as, "But I wouldn't use it because I don't see any need to, personally. I'm not saying other people wouldn't." Nonusers wanted to continue seeing their providers in person or talking on the phone, and users wanted to continue using the portal as is. A participant explained about possibly using the e-visit feature:

> *Usually, if I want to see the doctor, I want to see the doctor. And I know what it is why I'm going, and what it is I want to talk to them about. If it's just real*

> *simple, I can just e-mail him or call him. If I want to see my doctor, I undoubtedly have to make an appointment to see him. And if I want to see my doctor, I want to go see my doctor.*

Lack of Awareness of Functions Available

Intent to use was also influenced by participants' awareness of *My Health Manager* features and access to help using the website. Most participants did not know about the new features, and nonusers did not know about the basic features available via *My Health Manager* in general. As 1 nonuser participant stated, "You can see how I read on the computer, because I'd never seen that—make a—schedule an appointment. That wouldn't occur to me"; participants did not know what features were available or how to use them.

Discussion

Principal Findings

This study supports the growing literature suggesting many older patients, including those with MCC, are interested in using and are already using patient portals to help manage their health [22]. This is also the first study to use the TAM to qualitatively explore the connections between perceived usefulness, ease of use, and intent to use for a patient portal among older patients with MCC.

The TAM framework and supporting evidence [17,18] indicates several external variables influencing perceived ease of use with patient portals. Our study participants identified only 2 external variables: computer self-efficacy and anxiety. Specific patient portal user trainings offered in-person and/or on the Web may help older adults learn how to use the portal and when to use specific features [36]. Caregivers and family members are also helpful in reducing technology-specific anxiety [22,37], but more research is needed to inform portal design for shared access with caregivers [38].

Email, pharmacy, and medical lab result sections were popular and perceived as both useful and easy to use. This use behavior is consistent with other patient portal research in older populations [4,8]. These features are simple and quick while improving perceived patient-provider communication, satisfaction with access to health information, and fast medication management. Nonusers interested in the portal may be directed to these most popular, usable features. Research shows that once older adults are engaged in a technology, they tend to be high utilizers [23]. Therefore, promoting adoption of popular, easy-to-use features may foster patient satisfaction and further use of additional portal features. For example, promoting the email feature initially to encourage a patient to then try the portal pharmacy system.

Other features, particularly the Appointment Center, are difficult to use and do not offer perceived benefit to patients in this study. It is easier for patients to simply call to schedule appointments. There are also UI design issues related to small fonts and poor coloring, and negative UX influenced participants' intent to use the portal. These results align with the TAM and previous work suggesting that technology acceptance is determined by the perceived value and degree of burden. Older adults are unlikely to adopt burdensome technologies. Therefore, health systems should obtain ongoing UI and UX feedback from older adults with MCC when developing new tools and updates. The Department of Veteran Affairs implemented an ongoing feedback strategy that fostered adoption of their patient portal [39].

In terms of perceived usefulness, participants in this study suggested that patients *far away* from their providers would particularly benefit from the patient portal. However, older adults in rural areas are less likely to use patient portals [40], and internet use is lower among people in rural settings, especially among people with MCC [41]. Low internet and health technology use in rural communities is often attributed to limited access and awareness [42]. Recent improvements to broadband access [43] in rural communities may lead to increased portal adoption. However, more research is needed to determine best strategies for promoting portal engagement among older adults with MCC living in rural settings.

Participants indicated the portal would be helpful for *sicker* patients. Although we did not follow up to acquire a better definition of *sicker*, older patients with serious illness, owing to complex care needs, may benefit from portal use. There is some evidence substantiating increased portal adoption among older adults with worse health status [44]. A few studies indicate that people with cancer have positive perceptions of patient portals [45,46]. Although older adults with serious illness may be a target population for portal adoption, little is known about patient portal utility for patients with advanced or serious illness.

Preference for current methods is also a drawback to perceived usefulness and barrier to patients' intent to use. Participants value going to the clinic or pharmacy for physical activity and social engagement. With these values in mind, portal designers should consider adding functions that encourage older patients to *get out of the house* and connect personally with their providers. As portals advance, it is important to also respect the patients' need for a face-to-face connection with their providers. However, providers may consider using some face-to-face and phone-based time to encourage portal use [21]. Face-to-face and phone-based encouragement (eg, "Did you know you can schedule your next appointment in the convenience of your home online? Just go to My Health Manager") from providers and staff may increase patient awareness of beneficial features. In this study, health management tools and newer features were not used primarily owing to a lack of awareness.

In this sample, patient portals are not preferred by everyone, and other older adults with MCC may feel similarly. Usage varies greatly: some patients will never use the portal, other current users will continue to use only a few features, whereas another group will use every available option. Explicit nonusers appear to prefer human and face-to-face contact, which has previously been reported from a diverse sample of Kaiser patients [47]. Regardless of preference, technology-based health care interactions are increasing, and portal use may be expected. Addressing UI and UX challenges and promoting perceived benefits (improving commination, saving time, and access to personal information) may improve the intent to use patient portals among older adults with MCC.

Limitations

Although our study is an in-depth analysis of perceptions of older adults with MCC of a specific patient portal, the use behaviors and experience may differ across portal systems. While employing a descriptive qualitative approach for understanding portal UI and UX, intent to use, and use behavior of My Health Manager, we used only focus groups for data collection. This work would have benefited from inclusion of other data sources such as observation. We were unable to recruit as many portal nonusers resulting in limitations of our nonuser feedback. Although we recruited participants representing a wide age range, we did not maintain equal participation from each age group. The sample was also relatively well educated, middle income, and technology users, lacking specific input from underprivileged populations with less access to technology resources.

Conclusions

Older adults are interested in using patient portals and are already taking advantage of the features available to them. We have the opportunity to better engage older adults to use portals but need to pay close attention to key considerations promoting usefulness and ease of use. We recommend implementing portal user trainings, family and caregiver support, ongoing user feedback, and provider encouragement to improve intent to use and adoption among older adults with MCC.

Acknowledgments

The authors thank Dr Ted Palen for assistance with *My Health Manager*. This research was supported by funding from the National Institute on Aging (5T32AG044296). Dr Portz was supported by a career development award funded by the National Institute on Aging (K76AG059934).

Conflicts of Interest

None declared.

References

1. ONC: Office of the National Coordinator for Health Information Technology. What is a patient portal? URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2018-12-18] [WebCite Cache ID 74lVGyvCj]
2. ONC: Office of the National Coordinator for Health Information Technology. Are there difference types of personal health records (PHRs)? URL: https://www.healthit.gov/faq/are-there-different-types-personal-health-records-phrs [accessed 2018-12-18] [WebCite Cache ID 74lVwoHMC]
3. Coughlin SS, Prochaska JJ, Williams LB, Besenyi GM, Heboyan V, Goggans DS, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy 2017;10:33-40 [FREE Full text] [doi: 10.2147/RMHP.S130431] [Medline: 28435342]
4. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
5. 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]
6. 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;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
7. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
8. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
9. Frost and Sullivan. Trade NAVI. 2018. US Patient Portal Market for Hospitals and Physicians Overview and Outlook, 2012-2017 URL: http://tradenavi.net/CmsWeb/resource/attach/report/[688]U-S-%20Patient%20Portal%20Market%20for%20Hospitals%20and%20Physicians-pdf.pdf [accessed 2018-12-01] [WebCite Cache ID 74lYjIRuc]
10. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
11. Centers for Disease Control and Prevention. 2018. Healthy Aging - State of Aging and Health in America Report Internet URL: https://www.cdc.gov/aging/pdf/State-Aging-Health-in-America-2013.pdf [accessed 2018-07-12] [WebCite Cache ID 70rlwMs1W]
12. Federal Interagency Forum on Aging Related Statistics. Washington, DC; 2016. Older Americans 2016: Key Indicators of Well-Being Internet URL: https://agingstats.gov/docs/LatestReport/Older-Americans-2016-Key-Indicators-of-WellBeing.pdf [WebCite Cache ID 70rmCzXEh]
13. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth Enhanced Chronic Care Model: a theory derivation approach. J Med Internet Res 2015;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]

14. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015 Dec;18(6):3262-3273. [doi: 10.1111/hex.12316] [Medline: 25475371]
15. Pew Research Center. 2018. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/ [WebCite Cache ID 74nPx078E]
16. Pew Research Center. 2018. Tech Adoption Climbs Among Older Adults Internet URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2017/05/16170850/PI_2017.05.17_Older-Americans-Tech_FINAL.pdf [accessed 2018-07-13] [WebCite Cache ID 70tI2l9Ir]
17. 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 Apr 25;22(4):888-895. [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
18. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
19. Price MM, Pak R, Müller H, Stronge A. Older adults’ perceptions of usefulness of personal health records. Univ Access Inf Soc 2012 Mar 8;12(2):191-204. [doi: 10.1007/s10209-012-0275-y]
20. Nahm E, Sagherian K, Zhu S. Use of patient portals in older adults: a comparison of three samples. Stud Health Technol Inform 2016;225:354-358. [Medline: 27332221]
21. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito DA. 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]
22. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
23. Wildenbos GA, Peute L, Jaspers M. Facilitators and barriers of electronic health record patient portal adoption by older adults: a literature study. Stud Health Technol Inform 2017;235:308-312. [doi: 10.3233/978-1-61499-753-5-308] [Medline: 28423804]
24. Canziba E. Hands-On UX Design for Developers: Design, Prototype, and Implement Compelling User Experiences from Scratch. Birmingham, UK: Packt Publishing Ltd; 2018.
25. Fisk A, Rogers W, Charness N, Czaja S, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition. Boca Raton: CRC Press; 2009.
26. Greenberg AJ, Falisi AL, Finney RL, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 2;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
27. Redelmeier DA, Kraus NC. Patterns in patient access and utilization of online medical records: analysis of MyChart. J Med Internet Res 2018 Feb 6;20(2):e43 [FREE Full text] [doi: 10.2196/jmir.8372] [Medline: 29410386]
28. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
29. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
30. Groves S, Burns N, Jennifer G. The Practice Of Nursing Research: Appraisal, Synthesis, And Generation Of Evidence. St Louis, Missouri: Elsevier Saunders; 2019.
31. Nayar S, Stanley M. Qualitative Research Methodologies for Occupational Science and Therapy. London, UK: Routledge; 2016.
32. Bradshaw C, Atkinson S, Doody O. Employing a qualitative description approach in health care research. Glob Qual Nurs Res 2017;4:2333393617742282 [FREE Full text] [doi: 10.1177/2333393617742282] [Medline: 29204457]
33. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
34. Sharit J. The roles of health literacy, numeracy, and graph literacy on the usability of the VA's personal health record by veterans. J Usability Stud 2014;9(4):173-193 [FREE Full text]
35. Saldana J. The Coding Manual for Qualitative Researchers. Thousand Oaks, California: Sage Publications; 2009.
36. Sieck CJ, Hefner JL, Schnierle J, Florian H, Agarwal A, Rundell K, et al. he rules of engagement: perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 4;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]
37. Sarkar U, Bates DW. Care partners and online patient portals. J Am Med Assoc 2014;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
38. Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller DP, Smith D, et al. Insights into older adult patient concerns around the caregiver proxy portal use: qualitative interview study. J Med Internet Res 2018 Nov 2;20(11):e10524 [FREE Full text] [doi: 10.2196/10524] [Medline: 30389654]

39. Nazi KM, Turvey CL, Klein DM, Hogan TP. A decade of veteran voices: examining patient portal enhancements through the lens of user-centered design. J Med Internet Res 2018 Jul 10;20(7):e10413 [FREE Full text] [doi: 10.2196/10413] [Medline: 29991468]
40. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
41. Wang J, Bennett K, Probst J. Subdividing the digital divide: differences in internet access and use among rural residents with medical limitations. J Med Internet Res 2011;13(1):e25 [FREE Full text] [doi: 10.2196/jmir.1534] [Medline: 21371989]
42. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. J Rural Health 2018 Feb;34(Suppl 1):s30-s38. [doi: 10.1111/jrh.12228] [Medline: 28075508]
43. Federal Communications Commission. 2016. 2016 Broadband Progress Report Internet URL: https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2016-broadband-progress-report [accessed 2018-12-19] [WebCite Cache ID 74mvnXiHJ]
44. 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 2006:514-518 [FREE Full text] [doi: 10.7717/peerj.3268] [Medline: 17238394]
45. Zide M, Caswell K, Peterson E, Aberle DR, Bui AA, Arnold CW. Consumers' patient portal preferences and health literacy: a survey using crowdsourcing. JMIR Res Protoc 2016 Jun 8;5(2):e104. [doi: 10.2196/resprot.5122] [Medline: 27278634]
46. Alpert JM, Morris BB, Thomson MD, Matin K, Brown RF. Implications of patient portal transparency in oncology: qualitative interview study on the experiences of patients, oncologists, and medical informaticists. JMIR Cancer 2018 Mar 26;4(1):e5 [FREE Full text] [doi: 10.2196/cancer.8993] [Medline: 29581090]
47. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": Understanding Barriers to Portal Use among African Americans and Latinos. J Med Internet Res 2016 Dec 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]

Abbreviations

MCC: multiple chronic conditions
KPCO: Kaiser Permanente Colorado
TAM: Technology Acceptance Model
UI: user interface
UX: user experience

Edited by G Eysenbach; submitted 19.07.18; peer-reviewed by T Irizarry, C Jacob, J Hefner, T Risling; comments to author 08.10.18; revised version received 21.12.18; accepted 23.01.19; published 08.04.19.

Please cite as:
Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, Czaja S
Using the Technology Acceptance Model to Explore User Experience, Intent to Use, and Use Behavior of a Patient Portal Among Older Adults With Multiple Chronic Conditions: Descriptive Qualitative Study
J Med Internet Res 2019;21(4):e11604
URL: https://www.jmir.org/2019/4/e11604/
doi:10.2196/11604
PMID:30958272

Review

Patient Portals Facilitating Engagement With Inpatient Electronic Medical Records: A Systematic Review

Ronald Dendere[1*], PhD; Christine Slade[2*], PhD; Andrew Burton-Jones[3*], PhD; Clair Sullivan[1,4*], MBBS (Hons), MD, FACHI; Andrew Staib[5,6*], MBBS, FACHI; Monika Janda[1*], PhD

[1]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Woolloongabba, Australia
[2]Institute for Teaching and Learning Innovation, The University of Queensland, Brisbane, Australia
[3]School of Business, Faculty of Business, Economics and Law, The University of Queensland, Brisbane, Australia
[4]Metro North Hospital and Health Service, Department of Health, Queensland Government, Brisbane, Australia
[5]Metro South Hospital and Health Service, Department of Health, Queensland Government, Brisbane, Australia
[6]University of Queensland School of Medicine, Brisbane, Australia
[*]all authors contributed equally

Corresponding Author:
Monika Janda, PhD
Centre for Health Services Research
Faculty of Medicine
The University of Queensland
Level 2, Building 33, Princess Alexandra Hospital
Woolloongabba,
Australia
Phone: 61 7 3176 4569
Fax: 61 7 3365 1488
Email: m.janda@uq.edu.au

Abstract

Background: Engaging patients in the delivery of health care has the potential to improve health outcomes and patient satisfaction. Patient portals may enhance patient engagement by enabling patients to access their electronic medical records (EMRs) and facilitating secure patient-provider communication.

Objective: The aim of this study was to review literature describing patient portals tethered to an EMR in inpatient settings, their role in patient engagement, and their impact on health care delivery in order to identify factors and best practices for successful implementation of this technology and areas that require further research.

Methods: A systematic search for articles in the PubMed, CINAHL, and Embase databases was conducted using keywords associated with patient engagement, electronic health records, and patient portals and their respective subject headings in each database. Articles for inclusion were evaluated for quality using A Measurement Tool to Assess Systematic Reviews (AMSTAR) for systematic review articles and the Quality Assessment Tool for Studies with Diverse Designs for empirical studies. Included studies were categorized by their focus on input factors (eg, portal design), process factors (eg, portal use), and output factors (eg, benefits) and by the valence of their findings regarding patient portals (ie, positive, negative, or mixed).

Results: The systematic search identified 58 articles for inclusion. The inputs category was addressed by 40 articles, while the processes and outputs categories were addressed by 36 and 46 articles, respectively: 47 articles addressed multiple themes across the three categories, and 11 addressed only a single theme. Nineteen articles had high- to very high-quality, 21 had medium quality, and 18 had low- to very low-quality. Findings in the inputs category showed wide-ranging portal designs; patients' privacy concerns and lack of encouragement from providers were among portal adoption barriers while information access and patient-provider communication were among facilitators. Several methods were used to train portal users with varying success. In the processes category, sociodemographic characteristics and medical conditions of patients were predictors of portal use; some patients wanted unlimited access to their EMRs, personalized health education, and nonclinical information; and patients were keen to use portals for communicating with their health care teams. In the outputs category, some but not all studies found patient portals improved patient engagement; patients perceived some portal functions as inadequate but others as useful; patients and staff thought portals may improve patient care but could cause anxiety in some patients; and portals improved patient safety, adherence to medications, and patient-provider communication but had no impact on objective health outcomes.

Conclusions: While the evidence is currently immature, patient portals have demonstrated benefit by enabling the discovery of medical errors, improving adherence to medications, and providing patient-provider communication, etc. High-quality studies are needed to fully understand, improve, and evaluate their impact.

(J Med Internet Res 2019;21(4):e12779) doi:10.2196/12779

KEYWORDS

patient portal; electronic medical record; electronic health record; patient engagement; digital hospital

Introduction

The increasing adoption of electronic medical records (EMRs) by hospitals presents an opportunity for patients to access their clinical data and actively participate in their care via the EMR. Hospitals and other health care organizations can facilitate patient access to their EMR information through patient portals. Patient portals can provide secure, online access to personal health information [1] such as medication lists, laboratory results, immunizations, allergies, and discharge information [2]. They can also enable patient-provider communication using secure messaging, appointments and payment management, and prescription refill requests [2,3].

The increase in patient portal implementation is, in part, due to some preliminary evidence that they may improve patient engagement [4] and health outcomes such as medication adherence [5-10]. Government incentive programs and regulations also influenced some health care organizations to implement patient portals [11,12]. For example, in the United States, implementing patient portals was a way to meet the requirements for Meaningful Use, Stage 2, of the Healthcare Information Technology for Economic and Clinical Health Act [13].

Promoting patient involvement in health care delivery may lead to improved quality and safety of care [14,15] by enabling patients to spot and report errors in EMRs, for example [6]. Some patients recognize the role of patient portals in their health care, reporting satisfaction with the ability to communicate with their health care teams and perform tasks such as requesting prescription refills conveniently [3,16]. Portal use may reduce in-person visits, visits to emergency departments, and patient-provider telephone conversations [3,8-10,12,16]. Despite the potential of portals, already used in the ambulatory setting for some time, implementation in the inpatient setting has only recently gathered momentum [17-19]. The inpatient setting presents additional challenges for implementing patient portals [18,20]. Clinical conditions leading to hospitalization are often acute and the amount of medical information generated during this time can be extensive, which may overwhelm patients [20] and challenge information technology to rapidly display this information.

The aim of this study was to review literature describing patient portals tethered to an EMR in inpatient settings, their role in patient engagement, and their impact on health care delivery in order to identify factors and best practices for successful implementation of this technology and areas that require further research. Our review aims to inform researchers, health care organizations, and policymakers.

Methods

Search Strategy

The PubMed, CINAHL, and Embase databases were searched for articles published between 2005 and 2017 using keywords related to patient engagement, electronic health records, patient portals, and their associated subject headings in each database: the full search terms for each database are provided in Multimedia Appendix 1.

Study Selection and Quality Assessment

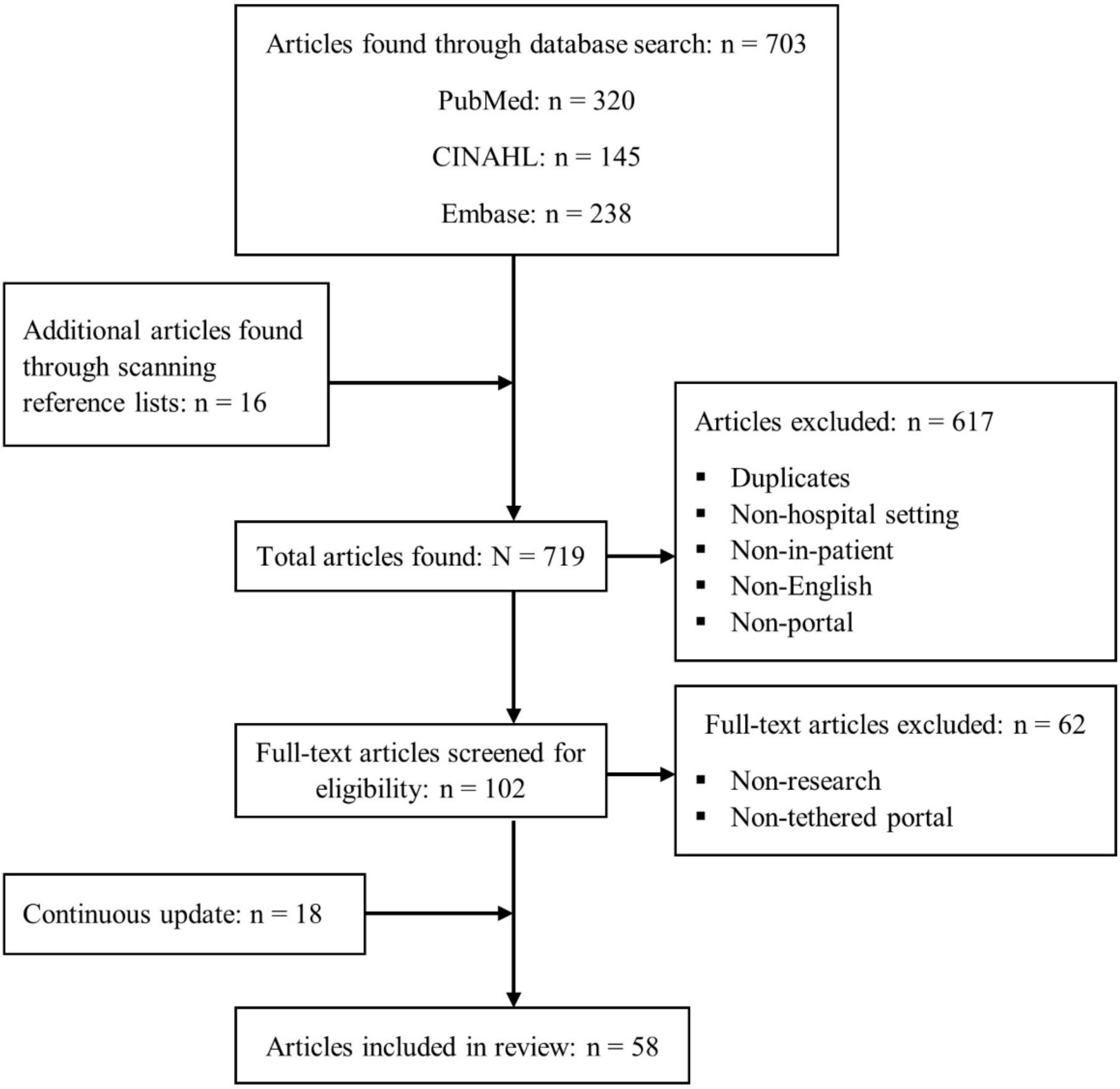
Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for the systematic search and selection process. Inclusion criteria for articles were (1) written in English, (2) hospital inpatient setting, and (3) patient portals tethered to a hospital EMR. The initial combined database search produced 703 articles, and an additional 16 were identified by scanning their reference lists. After eliminating duplicates, the article abstracts were independently reviewed by three authors to identify articles that did not meet the inclusion criteria. This led to 617 articles being excluded. Full-text screening was conducted for the remaining 102 articles, leading to the identification of 62 articles that did not meet the inclusion criteria. At each stage, the authors met to reconcile, by consensus, any disagreements about article inclusion. An independent coder also coded the 102 articles for inclusion/exclusion using our criteria, and interrater agreement was high (Cohen kappa=.75). In cases of disagreement, we opted to include the article if it addressed a potentially important policy issue (eg, privacy issues, rural/urban divide). To ensure we included as many up-to-date papers as possible, we periodically conducted database searches for new articles after the initial search. This step was performed by just one author because by this stage, the authors had established a well-developed understanding of the inclusion/exclusion criteria. This periodic update, up to August 2018, identified another 18 articles for inclusion.

The included articles were assessed for quality. Two authors independently scored each article’s quality using the most recent version of A Measurement Tool to Assess Systematic Reviews (AMSTAR 2) [21] for review articles and the Quality Assessment Tool for Studies with Diverse Designs (QATSDD) [22] for qualitative, quantitative, and mixed empirical studies. We classified the overall AMSTAR 2– or QATSDD-derived score for each paper on a 5-point scale (very low, low, medium, high, very high), thus establishing the qualities of all articles on one scale. The ratings by the authors and the independent coder were highly correlated (*r*=.81). A third coauthor then independently reviewed and reconciled the scores. We assigned valence ratings to each article to characterize the overall findings in each article regarding patient portals as positive, negative, or mixed [23].

Data Analysis

We analyzed the information extracted from the included articles by categorizing the themes related to the implementation of patient portals into inputs, processes, and outputs. The inputs are the material (eg, hardware and software) and nonmaterial (eg, leadership) components that facilitate or impair the establishment or use of the portal. Processes include the interactions of the users with the portal. Outputs comprise the results of the implementation or the use of the portal. Through the analysis, we identified 14 themes within these three categories, shown in Textbox 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of the search and selection process.



Textbox 1. Description of the themes identified in the implementation of inpatient portals.

- Input themes
    - Portal design: umbrella term for all design-related aspects of the portal including portal interface, content, features, and functions
    - Usability: extent to which a patient portal has the property of being able to be used by patients, caregivers, and health care teams to enhance patient engagement with effectiveness, efficiency, and satisfaction
    - Barriers: factors that hinder widespread adoption or portal use
    - Facilitators: factors that motivate or enable users to sign up for or actively use a portal
    - User training: equipping patients and health care team members with the necessary skills and knowledge to effectively use a portal
    - Organizational factors: culture of a health care organization; decisions and actions it takes when an initial consideration is made to implement a patient portal
- Process themes
    - Adoption: from a patient perspective, adoption is the registration for a portal account; from a health care provider perspective, portal adoption refers to acceptance and promotion of the portal [16]
    - Use: active engagement and continued use after signing up for a portal [4]
    - Information: all aspects associated with providing patients with clinical and nonclinical information via a portal
    - Communication: all aspects associated with portal-based patient-provider communication
- Output themes
    - Patient engagement: active involvement of patients in their own health care
    - User perceptions: thoughts, feelings, and opinions of patients, caregivers, and health care team members about their experiences with a patient portal
    - Health outcomes: impact of patient portals on clinical indicators
    - Benefits: value provided by patient portals to the health care delivery process

Results

Overview: Search Strategy, Study Selection, and Quality Assessment

Details of the 58 articles included in this review are given in Multimedia Appendices 2 and 3. The majority of the articles (41) are indexed in all three databases; 13 articles are indexed in two databases, and only 4 are in indexed in a single database. The PubMed and Embase databases each index 56 articles while CINAHL indexes 41 articles. Nineteen articles described qualitative studies, 18 quantitative studies, and 9 mixed method studies, while 12 were reviews. Our quality assessment placed 19 articles in the high- or very high–quality categories, 21 in the medium category, and 18 in the low- or very low–quality categories. Twenty-nine articles were assigned mixed valence, 16 articles were assigned positive valence, and 2 were assigned negative valence. Valence was not applicable or could not be drawn in 11 articles (ie, those that examined portals but that were not focused on evaluating them in any way).

There was a spread of articles addressing themes over the inputs, processes and outputs categories (see Multimedia Appendix 3). However, only one-third of the articles (19/58) addressed the full spectrum of categories.

Textbox 2 shows that there was corroboration among studies for some findings (eg, information) but contradicting findings in some themes (eg, health outcomes). We did not observe trends in the findings that were related to study quality or design (ie, there was no association between study quality or study design and reported findings).

Textbox 2. Summary of findings in the reviewed literature.

- Inputs
    - Portal design
        - Portals were designed using an iterative approach [6,16,24]
        - Integrated infobuttons linked users with reliable sources of medical information [19,25-27]
        - Artificial intelligence was used to enhance portal designs [28]
        - Portals can be designed for specific diseases or medical conditions [29]
        - Patients requested electronic games within portals [18,30] and functionality to control caregivers' access to their electronic medical records (EMRs) and receive notifications when their EMRs had been viewed by a caregiver [31]
    - Usability
        - Some participants had difficulties using patient portals mainly because of complex portal interfaces [19,24,32-34]
        - Some patients found patient portals easy to use [17,30,35-37]
    - Barriers
        - Lack of appropriate training [33]
        - Doubt of the portal's usefulness [38]
        - Lost passwords [38,39]
        - Difficulties in using portals [16,33,34]
        - Anxiety associated with viewing personal medical information [38]
        - Data security and privacy concerns [34,40]
        - Lack of encouragement from providers [3,40]
    - Facilitators
        - Access to information [18,19,24,30,36,40-42]
        - Patient-provider communication [26,34,40,43]
        - Record-keeping [36,42,43]
        - Provider encouragement [40]
    - User training
        - Patients were trained using videos and reading material [24,33]
        - Health care teams were trained using verbal instructions and hands-on sessions [5,30,44,45]
    - Organizational factors
        - Rural hospitals were more likely to report costs and obtaining staff cooperation as barriers to health information technology (HIT; including patient portals) adoption than urban hospitals [46]
        - Small hospitals were more likely to report cost-related barriers than large hospitals [46]
        - Leaderships had crucial roles in the implementation of patient portals, working closely with developers in system design, developing policies to guide user training, and integrating portals into clinical workflows [33]
        - Implementation of patient portals varied across organizations due to different interpretations of government legislations by the health care organizations [47]
- Processes
    - Adoption
        - Portal use was higher among white patients than other racial groups, younger patients than older patients, female patients than male patients, and high-income than low-income patients [4,6,33,48]
    - Use
        - Portal use was higher among patients with greater disease severity [4]
        - Patients were less inclined to use a portal when they were seriously ill, in intense pain,

    or soon after undergoing multiple tests or procedures [24]

  - Information
    - Patients wanted timely and comprehensive access to their medical information [6,18,24,26,30]
    - Some patients preferred to have access to their entire EMR, including doctors' notes [24,41]
    - Patients wanted personalized information tailored to their conditions and needs [6,24,35,49]
    - Patients requested clinical unit maps, meal menus [30], and short biographies of their health care team members [19,24]
  - Communication
    - Patients and caregivers expressed interest in using portals to communicate with health care staff [6,24,30,36] but not many actually used this feature [30,44]
    - Patients used portal messaging to request information, communicate needs and concerns, contribute to care coordination, offer feedback [26], compliment health care staff, and express gratitude [30,45]
    - Some patients wanted an option to send messages to specific staff members and an indication of whether a message had been read and when to expect a response [33]
    - Ethnicity, age, and gender were associated with portal-based communication [7,16,33,48]
- Outputs
  - Patient engagement
    - Some portals did not significantly improve patient engagement [17,50] but others did [6,41,43,51,52]
    - Patients in some studies reported that portals enabled better engagement in their own care [19,24,30]
  - User perceptions
    - Some patients felt that portals did not adequately fulfill their information needs [6,24,29,42]
    - Patients associated unrestricted access to their EMRs with empowerment and a sense of control [24,36,42,53]
    - Some patients and health care staff had concerns that unrestricted access to sensitive information may cause anxiety and more questions for health care staff [24,41,49]
    - Health care teams had preimplementation concerns about disruptions to workflows and potential for large volumes of patient messages, but such concerns did not materialize [45]
    - Patients and staff thought that the messaging feature of patient portals was important for patient care [17,24,26,30]
    - Patients who used a disease-specific portal were more satisfied than those who used a generic portal [29]
    - Patients and health care staff agreed that patient portals helped to improve patient care [5,30,44]
  - Health outcomes
    - Significant association between portal use and health outcomes was not observed in some studies [2,4,20] but was observed in others [54]
    - Patient portals facilitated discovery of EMR errors by patients [24,26,30,54-56]
  - Benefits
    - Improved adherence to medication [3,9,39,43,52,57]
    - Improved patient satisfaction [3,29,30,39,45]
    - Enhanced patient-provider communication [6,30,40,43,52]
    - Improved patient safety [6,26,52,55]
    - Reduced patient uncertainty and anxiety [3,24,36]
    - Increased patient engagement [6,19,24,30,41,43,45,51,52]

Patient Portal Inputs

Forty articles addressed themes in the inputs category. We identified 22 articles that addressed portal design. In an iterative design approach, feedback from patients, including requests for electronic games and other functions, was used to refine designs. Enhancements included links to medical education, artificial intelligence techniques, and disease-specific design.

As shown in Textbox 2, poor portal designs caused usability difficulties for some patients. Those difficulties were among the barriers to portal use. Textbox 2 also shows that more barriers to portal use than facilitators were identified in the reviewed literature. Various methods were used to train patients and staff to use the portals. However, training methods were not optimal [5,33,44]. For example, in one study, patient training was delivered via an 11-minute video that was not well received, with less than a third (26.3%) watching the entire video [33]. In another study among health care staff, doctors had the lowest confidence in a patient portal, and they doubted their training was sufficient to allow them to effectively use it [5].

Organizational factors (leadership, staff support, and key decisions, etc) was the least addressed theme. However, the findings summarized in Textbox 2 indicate that organizational factors are likely the most crucial in determining whether or how patient portals are implemented.

Patient Portal Processes

Thirty-six articles addressed themes categorized as processes. Five articles addressed portal adoption while use was addressed in 18 articles (see Multimedia Appendix 3). Sociodemographic characteristics of patients such as race, age, gender, level of education, and social status were predictors of both portal adoption and use [4,6,33,48]. In addition to sociodemographic factors, there was higher portal use among patients with greater disease severity (eg, advanced cancer) [4], but at any given time, patient condition influenced portal use as described in Textbox 2.

Seventeen articles addressed information and 20 addressed communication. The articles suggested that patients wanted unlimited access to their EMRs, medical education in layman's terms [24,41], and nonclinical information [19,24,30].

Patient-provider communication, usually in the form of secure messages, is a key feature of inpatient portals [26]. Despite expressions of interest in this feature, actual use was low [30,44]; in a pediatric study with 296 parents participating, only about 6% sent messages to health care teams via the inpatient portal [30]. Most portals enabled patients to send messages to a single mailbox that was accessed by staff members on duty, but some patients wanted to communicate with specific staff members [33]. Similar to adoption and use, sociodemographic factors were predictors of which patients used the messaging feature (see Textbox 2).

Patient Portal Outputs

The outputs category was addressed by 46 articles, and 24 articles addressed patient engagement. Results of patient engagement were mixed: portals in some studies did not cause statistically significant improvement, but patients in other studies reported that portals enabled better engagement in their care.

User perceptions was the most commonly addressed theme across all categories (35 articles), and Textbox 2 shows that users perceived some functions of portals (eg, access to information) as inadequate but perceived other functions (eg, communication) as very useful. Despite dissatisfaction and concerns with some aspects of the portals, users' perceptions of the patient portal concept as a whole were mostly positive. For example, in one study, 90% of participants reported overall satisfaction with the portal, 89% thought that the portal reduced errors, and 94% agreed that the portal improved care delivery [30]. In another study, 84% of participants described a patient portal as useful and 90% reported that they would recommend it to their peers [32]. Health care staff appreciated the reasons for implementing patient portals and their own roles and responsibilities in the process, but they stated that they would like to receive sufficient training [5].

Twenty articles addressed the association between patient portals and health outcomes, such as medical errors, readmissions, and mortality. Results were mixed as some studies did not show positive associations between portal use and health outcomes [2,4,20]. For example, a retrospective study of cancer patients found no association between portal adoption or use and adverse events [4] and another retrospective study found no association between portal use and 30-day readmission or inpatient mortality [20]. However, in a third study, active portal use significantly improved glycated hemoglobin levels [54]. Importantly, patient portals in several studies facilitated discovery of EMR errors by patients, particularly medication errors [24,26,30,54-56].

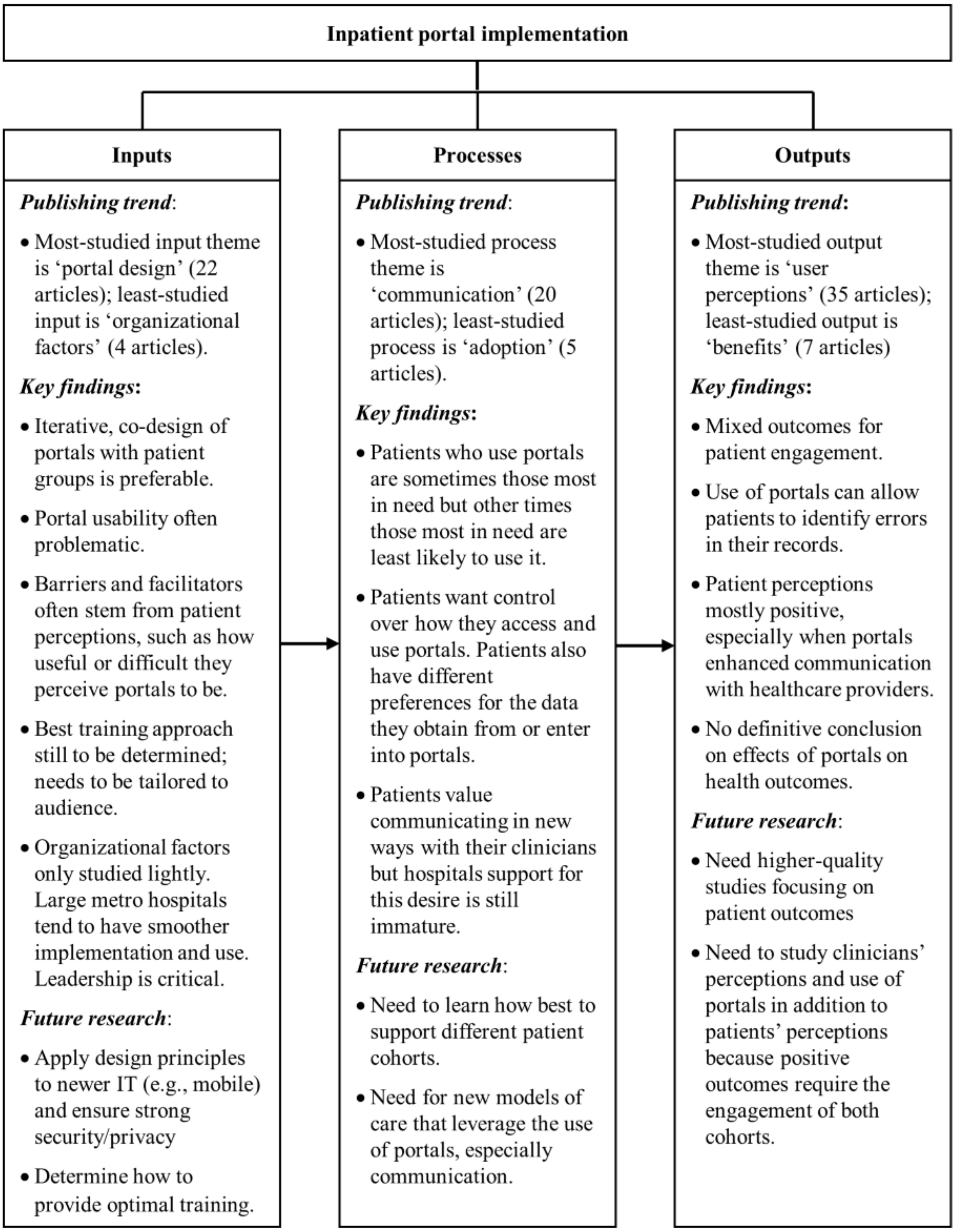
Finally, 7 articles addressed the benefits of patient portals, and Textbox 2 shows that portals offers a wide range of benefits for patient care.

Discussion

Principal Findings

This systematic review examined 58 articles studying inpatient portals. Although there was overlap in the themes reflected in these studies, there was also significant variation in the setting, patient population, software, outcomes assessed, and study methodology, making it hard to come to a definitive conclusion on whether inpatient portals are beneficial. This is further shown by the higher number of included studies that were judged to relay mixed valence than those with positive valence. However, more studies relayed positive than negative valence, showing that patient portals may be beneficial for health care. We discuss below the patient portal input, process and output factors that contribute to this assessment, and the areas of research that need focus in order to improve patient care. Figure 2 provides a visual summary.

Figure 2. Conceptual framework summarizing the findings and key areas for future research.



Inpatient portal implementation

Inputs

Publishing trend:

- Most-studied input theme is ‘portal design’ (22 articles); least-studied input is ‘organizational factors’ (4 articles).

Key findings:

- Iterative, co-design of portals with patient groups is preferable.
- Portal usability often problematic.
- Barriers and facilitators often stem from patient perceptions, such as how useful or difficult they perceive portals to be.
- Best training approach still to be determined; needs to be tailored to audience.
- Organizational factors only studied lightly. Large metro hospitals tend to have smoother implementation and use. Leadership is critical.

Future research:

- Apply design principles to newer IT (e.g., mobile) and ensure strong security/privacy
- Determine how to provide optimal training.

Processes

Publishing trend:

- Most-studied process theme is ‘communication’ (20 articles); least-studied process is ‘adoption’ (5 articles).

Key findings:

- Patients who use portals are sometimes those most in need but other times those most in need are least likely to use it.
- Patients want control over how they access and use portals. Patients also have different preferences for the data they obtain from or enter into portals.
- Patients value communicating in new ways with their clinicians but hospitals support for this desire is still immature.

Future research:

- Need to learn how best to support different patient cohorts.
- Need for new models of care that leverage the use of portals, especially communication.

Outputs

Publishing trend:

- Most-studied output theme is ‘user perceptions’ (35 articles); least-studied output is ‘benefits’ (7 articles)

Key findings:

- Mixed outcomes for patient engagement.
- Use of portals can allow patients to identify errors in their records.
- Patient perceptions mostly positive, especially when portals enhanced communication with healthcare providers.
- No definitive conclusion on effects of portals on health outcomes.

Future research:

- Need higher-quality studies focusing on patient outcomes
- Need to study clinicians’ perceptions and use of portals in addition to patients’ perceptions because positive outcomes require the engagement of both cohorts.

Patient Portal Inputs

When assessing common themes within the inputs category, portal design, usability, and barriers were more widely covered than user training and organizational factors. Many articles showed that good portal design is crucial for usability and adoption by patients. While most studies involved users in the design process, more needs to be done to overcome design-related barriers, particularly for people with low health literacy [6,19,24]. Increasing the use of enabling technology such as voice-commanded digital assistants, artificial intelligence, and natural language processing could make the systems cater to a wider range of audiences. Designing platforms so that they can be displayed on users’ personal mobile devices could be another enabler of portal adoption.

Patients’ concerns regarding privacy and security of their medical information [34,40] are particularly relevant in cases where caregivers need access to patients’ EMRs. Discussion of privacy issues in cases of caregiver access to patient medical information in the literature is limited. In the absence of formal policy for caregiver access, patients may opt to disclose personal log-in details to their caregivers, which is not recommended [26].

The results of user training showed that reliance on a one-size-fits-all approach may not be effective for educating users because of varying preferences. Training could be enhanced by providing information that directly addresses common patient concerns (eg, information security) and health care staff concerns (eg, workflow changes).

Information on the best timing to deliver training material to patients is lacking in the literature. While staff training can be scheduled ahead of deploying a patient portal, patients may only encounter a portal upon hospitalization. Furthermore, the severity of their condition may limit their ability to focus on or understand the training material. Innovation may be needed to inform patients about portal services prior to hospitalization. For example, patients could sign up for outpatient portals which would be similar in design to the inpatient portals to ensure seamless transition between the two portals. Hospitals could also collaborate with medical insurance providers to make training material available to potential patients already signing up for hospital insurance.

Patient Portal Processes

Associations between patient sociodemographic characteristics and portal use [4,6,33,48] indicate that patients who are most vulnerable (eg, those with low health literacy or seriously ill) would be least likely to benefit from patient portals. Recruitment and participation bias when engaging participants in user testing mean that actual use by patients may not reflect the use in the initial testing phases. Therefore, health care organizations may need to conduct multiple studies to iteratively address factors that influence portal use within the communities they serve.

The varying preference among patients for level of access to EMRs [6,18,24,26,30] and staff concerns with unlimited patient access to sensitive information (information that could cause anxiety for patients) [24,41,49] present challenges for health care organizations. Health care organizations could fulfill these diverse information needs by flagging sensitive information and warning patients that despite having access to the information, they can opt to receive it in person from a health care team member. Some hospitals already have guidelines for releasing sensitive information to outpatient portals [26], and those guidelines could be adapted to the inpatient setting.

Health care organizations may view patient requests of nonclinical information and functions such as electronic games [18,19,24,30] as fringe requests that could raise developmental costs. However, granting these requests may improve portal adoption and use and overall patient satisfaction. Also, patients have requested background information on their health care staff, but staff perceptions about the disclosure of such information to their patients is not known [19,24]. Further research is needed to uncover staff perceptions in this regard and explore ways of implementing this service.

Portal-based patient-provider communication is potentially beneficial but may also be disruptive [26]. Although staff concerns about potential interruptions caused by constant patient messaging did not materialize [45], in practice, staff members may be overwhelmed by messages at any time. Structured messaging may be a solution that ensures patients communicate only important and relevant information [26]. If health care organizations decide to enforce structured messaging, they should prioritize patient safety and therefore avoid restrictions that could prevent patients reporting genuine concerns.

Patient Portal Outputs

The literature explored some, but not all, potential outputs of patient portal implementations. Most of the studies assessed implementation of patient portals using interim outcomes such as user perceptions, and few studies addressed important objective outcomes such as length of stay, morbidity, or mortality [2,4,20,54].

Some studies showed no association between portal use and health outcomes such as readmission, adverse events, or mortality [2,4,20]. However, a number of those studies drew their conclusions from retrospective analysis of portal adoption and use data only [4,20]—adoption and frequency of use alone do not provide sufficient information about effective portal use, which may affect outcomes.

Increasing patient engagement is a goal of patient portals, but the engagement of health care staff is also important since they are likely to be approached by the patients with portal-related queries [5,44]. Nurses and doctors should have sufficient knowledge to answer basic questions or appropriately escalate complex questions (eg, to information technology support). Also, nurses may be required to respond to patient-generated messages within the portal. Similarly, doctors' perceptions of portals are also important, as they may use a portal to communicate with patients [5] and therefore need to be confident with its functions. Theories in health care information technology suggest that user perceptions can predict the acceptance and use of new technologies [58-60]. Therefore, it is essential for hospitals to ensure positive staff attitudes toward patient portals through effective staff training and technical support and incorporating staff needs in portal design and workflows.

Several studies reported that patient portals facilitate patient discovery of errors in EMRs [24,26,30,54-56]. Discussion in the literature of how patients could notify health care providers of such errors is lacking. Further research is needed to establish how patient-discovered errors are reported and to identify optimal reporting methods.

Research in the evaluation of patient portals is also currently limited. Standardized evaluation frameworks and measures are needed to enable better comparisons of patient portal implementation and outcomes in the future.

Limitations

While an extensive search was undertaken, the majority of the included studies were conducted at single locations, used outcome measures that were not comparable to those used in other studies, and had small sample sizes. That means results of those studies may not be generalizable to other population groups. Also, a number of the studies were conducted in controlled settings, such as closed-door observations and interviews, which would not be representative of hospital settings. The absence of standardized evaluation tools means the results could not be compared or synthesized, and we were thus limited to providing a descriptive summary of findings only. Most studies that addressed user perceptions or patient-reported results depended on the opinions of those who completed end-of-study questionnaires or interviews; such results could be biased as they may lack feedback from participants who felt uneasy about giving negative feedback. Finally, our conceptual framework enabled us to group the findings gleaned from the included articles, although there was some overlap in the categories caused by interdependence in some of the themes.

Conclusion

The review results suggest that the available evidence for inpatient portals is currently immature. Standardized outcomes assessment and more high-quality studies with objective outcomes (length of stay, mortality, and morbidity) are required to fully understand the impact of such portals.

Acknowledgments

This study benefited from the support of the Australian Research Council (ARC FT130100942), the National Health and Medical Research Council Translating Research into Practice Fellowship (APP 1151021), and the University of Queensland.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms for the PubMed, CINAHL, and Embase databases.

[PDF File (Adobe PDF File), 24KB - jmir_v21i4e12779_app1.pdf]

Multimedia Appendix 2

Detailed summary of each article selected for inclusion in the review.

[PDF File (Adobe PDF File), 122KB - jmir_v21i4e12779_app2.pdf]

Multimedia Appendix 3

Themes addressed by each article.

[PDF File (Adobe PDF File), 107KB - jmir_v21i4e12779_app3.pdf]

References

1. Office of the National Coordinator for Health Information Technology. 2017. What is a patient portal? URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2019-03-12] [WebCite Cache ID 73H6rv9V7]
2. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501. [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
3. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
4. Aljabri D, Dumitrascu A, Burton MC, White L, Khan M, Xirasagar S, et al. Patient portal adoption and use by hospitalized cancer patients: a retrospective study of its impact on adverse events, utilization, and patient satisfaction. BMC Med Inform Decis Mak 2018 Jul 27;18(1):70 [FREE Full text] [doi: 10.1186/s12911-018-0644-4] [Medline: 30053809]
5. Hefner JL, Sieck CJ, Walker DM, Huerta TR, McAlearney AS. System-wide inpatient portal implementation: survey of health care team perceptions. JMIR Med Inform 2017 Sep 14;5(3):e31 [FREE Full text] [doi: 10.2196/medinform.7707] [Medline: 28912115]
6. Kelly MM, Coller RJ, Hoonakker PL. Inpatient portals for hospitalized patients and caregivers: a systematic review. J Hosp Med 2018 Dec 01;13(6):405-412. [doi: 10.12788/jhm.2894] [Medline: 29261819]
7. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? A systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
8. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]

9. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):325-327 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
10. 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]
11. Reed P, Conrad DA, Hernandez SE, Watts C, Marcus-Smith M. Innovation in patient-centered care: lessons from a qualitative study of innovative health care organizations in Washington State. BMC Fam Pract 2012 Dec 14;13:120 [FREE Full text] [doi: 10.1186/1471-2296-13-120] [Medline: 23241305]
12. Australian Commission on Safety and Quality in Healthcare. 2018. Impact of digital health on the safety and quality of health care URL: https://tinyurl.com/y9puca9o [accessed 2018-10-19] [WebCite Cache ID 73H8Gzhzy]
13. National Learning Consortium. 2013. How to optimize patient portals for patient engagement and meet meaningful use requirement URL: https://www.healthit.gov/sites/default/files/nlc_how_to_optimizepatientportals_for_patientengagement.pdf [accessed 2019-03-12] [WebCite Cache ID 73HANVXfP]
14. Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, et al. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. BMJ 2001 Oct 20;323(7318):908-911 [FREE Full text] [Medline: 11668137]
15. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract 2000 Sep;49(9):796-804. [Medline: 11032203]
16. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
17. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy SG, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016 Jan;23(1):159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
18. Prey JE, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc 2014;21(4):742-750 [FREE Full text] [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]
19. Vawdrey DK, Wilcox LG, Collins SA, Bakken S, Feiner S, Boyer A, et al. A tablet computer application for patients to participate in their hospital care. AMIA Annu Symp Proc 2011;2011:1428-1435 [FREE Full text] [Medline: 22195206]
20. Dumitrascu AG, Burton MC, Dawson NL, Thomas CS, Nordan LM, Greig HE, et al. Patient portal use and hospital outcomes. J Am Med Inform Assoc 2018 Apr 01;25(4):447-453. [doi: 10.1093/jamia/ocx149] [Medline: 29300961]
21. 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. BMJ 2017 Dec 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
22. Sirriyeh R, Lawton R, Gardner P, Armitage G. Reviewing studies with diverse designs: the development and evaluation of a new tool. J Eval Clin Pract 2012 Aug;18(4):746-752. [doi: 10.1111/j.1365-2753.2011.01662.x] [Medline: 21410846]
23. de Matos CA, Rossi CAV. Word-of-mouth communications in marketing: a meta-analytic review of the antecedents and moderators. J Acad Mark Sci 2008 Sep 23;36(4):578-596. [doi: 10.1007/s11747-008-0121-1]
24. Woollen J, Prey J, Wilcox L, Sackeim A, Restaino S, Raza ST, et al. Patient experiences using an inpatient personal health record. Appl Clin Inform 2016;7(2):446-460 [FREE Full text] [doi: 10.4338/ACI-2015-10-RA-0130] [Medline: 27437053]
25. Borbolla D, Del Fiol G, Taliercio V, Otero C, Campos F, Martinez M, et al. Integrating personalized health information from MedlinePlus in a patient portal. Stud Health Technol Inform 2014;205:348-352. [Medline: 25160204]
26. Grossman LV, Choi SW, Collins S, Dykes PC, O'Leary KJ, Rizer M, et al. Implementation of acute care patient portals: recommendations on utility and use from six early adopters. J Am Med Inform Assoc 2018 Apr 01;25(4):370-379. [doi: 10.1093/jamia/ocx074] [Medline: 29040634]
27. Lister Hill National Center for Biomedical Communications, NLM. 2016. Infobuttons URL: https://lhncbc.nlm.nih.gov/project/infobuttons [accessed 2018-10-19] [WebCite Cache ID 73HFl8O5X]
28. Arnold CW, McNamara M, El-Saden S, Chen S, Taira RK, Bui AAT. Imaging informatics for consumer health: towards a radiology patient portal. J Am Med Inform Assoc 2013;20(6):1028-1036 [FREE Full text] [doi: 10.1136/amiajnl-2012-001457] [Medline: 23739614]
29. Lee J, Kim JGB, Jin M, Ahn K, Kim B, Kim S, et al. Beneficial effects of two types of personal health record services connected with electronic medical records within the hospital setting. Comput Inform Nurs 2017 Nov;35(11):574-581. [doi: 10.1097/CIN.0000000000000362] [Medline: 28548973]
30. Kelly MM, Hoonakker PLT, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Dec;24(1):153-161. [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
31. Caine K, Kohn S, Lawrence C, Hanania R, Meslin EM, Tierney WM. Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews. J Gen Intern Med 2015 Jan;30 Suppl 1:S7-S16 [FREE Full text] [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]

32. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
33. Yen P, Walker DM, Smith JMG, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. Int J Med Inform 2018 Dec;110:10-18. [doi: 10.1016/j.ijmedinf.2017.11.007] [Medline: 29331248]
34. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
35. Runaas L, Bischoff E, Hoodin F, Kentor R, Ostarello L, Seyedsalehi S, et al. A novel health informatics tool to improve caregiver activation: findings from pediatric BMT in a hospital-based setting. Blood 2016;128(22):2382 [FREE Full text]
36. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
37. 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;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
38. Hazara AM, Bhandari S. Barriers to patient participation in a self-management and education website Renal PatientView: a questionnaire-based study of inactive users. Int J Med Inform 2016 Mar;87:10-14. [doi: 10.1016/j.ijmedinf.2015.12.004] [Medline: 26806707]
39. 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]
40. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
41. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
42. Rexhepi H, Åhlfeldt R, Cajander A, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: a qualitative study. Health Informatics J 2016 Jul 19;24(2):115-124. [doi: 10.1177/1460458216658778] [Medline: 27440056]
43. Rathert C, Mittler JN, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: what does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]
44. Hefner JL, Sieck CJ, McAlearney AS. Training to optimize collaborative use of an inpatient portal. Appl Clin Inform 2018 Jul;9(3):558-564. [doi: 10.1055/s-0038-1666993] [Medline: 30045386]
45. Kelly MM, Dean SM, Carayon P, Wetterneck TB, Hoonakker PLT. Healthcare team perceptions of a portal for parents of hospitalized children before and after implementation. Appl Clin Inform 2017 Dec 15;8(1):265-278 [FREE Full text] [doi: 10.4338/ACI-2016-11-RA-0194] [Medline: 28293685]
46. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
47. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy and Technology 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]
48. Mikles SP, Mielenz TJ. Characteristics of electronic patient-provider messaging system utilisation in an urban health care organisation. J Innov Health Inform 2014 Dec 18;22(1):214-221 [FREE Full text] [doi: 10.14236/jhi.v22i1.75] [Medline: 25924551]
49. Walker DM, Menser T, Yen P, McAlearney AS. Optimizing the user experience: identifying opportunities to improve use of an inpatient portal. Appl Clin Inform 2018 Dec;9(1):105-113. [doi: 10.1055/s-0037-1621732] [Medline: 29444536]
50. Ralston JD, Silverberg MJ, Grothaus L, Leyden WA, Ross T, Stewart C, et al. Use of web-based shared medical records among patients with HIV. Am J Manag Care 2013 Apr 01;19(4):e114-e124 [FREE Full text] [Medline: 23725449]
51. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with ehealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
52. van der Vaart R, Drossaert CH, Taal E, van de Laar MA. Giving rheumatology patients online home access to their electronic medical record (EMR): advantages, drawbacks and preconditions according to care providers. Rheumatol Int 2013 Sep;33(9):2405-2410 [FREE Full text] [doi: 10.1007/s00296-012-2408-2] [Medline: 22453527]
53. Davis GT, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014;21(4):737-741. [doi: 10.1136/amiajnl-2013-002239] [Medline: 24154835]
54. Toscos T, Daley C, Heral L, Doshi R, Chen Y, Eckert GJ, et al. Impact of electronic personal health record use on engagement and intermediate health outcomes among cardiac patients: a quasi-experimental study. J Am Med Inform Assoc 2016 Jan;23(1):119-128. [doi: 10.1093/jamia/ocv164] [Medline: 26912538]

55. 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):e157-e162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
56. Wright A, Feblowitz J, Maloney FL, Henkin S, Ramelson H, Feltman J, et al. Increasing patient engagement: patients' responses to viewing problem lists online. Appl Clin Inform 2014;5(4):930-942. [doi: 10.4338/ACI-2014-07-RA-0057] [Medline: 25589908]
57. Saberi P, Catz SL, Leyden WA, Stewart C, Ralston JD, Horberg MA, et al. Antiretroviral therapy adherence and use of an electronic shared medical record among people living with HIV. AIDS Behav 2015 Jun;19 Suppl 2:177-185 [FREE Full text] [doi: 10.1007/s10461-014-0982-x] [Medline: 25572829]
58. Carayon P, Cartmill R, Blosky MA, Brown R, Hackenberg M, Hoonakker P, et al. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc 2011;18(6):812-819 [FREE Full text] [doi: 10.1136/amiajnl-2010-000018] [Medline: 21697291]
59. Karsh B. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
60. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
EMR: electronic medical record
HIT: health information technology
QATSDD: Quality Assessment Tool for Studies With Diverse Designs

Edited by G Eysenbach; submitted 18.11.18; peer-reviewed by B Xie, A McAlearney, G Strudwick, T Irizarry; comments to author 27.12.18; revised version received 30.01.19; accepted 17.02.19; published 11.04.19.

Please cite as:
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;21(4):e12779
URL: http://www.jmir.org/2019/4/e12779/
doi:10.2196/12779
PMID:30973347

Original Paper

Racial, Ethnic, and Socioeconomic Disparities in Web-Based Patient Portal Usage Among Kidney and Liver Transplant Recipients: Cross-Sectional Study

Joel Wedd[1,2], MD, MPH; Mohua Basu[1], MPH; Laura M Curtis[3], MS; Kayla Smith[1], BS; Denise J Lo[1], MD; Marina Serper[4], MD; Michael S Wolf[3], MD, PhD; Ruth Parker[1], MD; Rachel E Patzer[1], PhD, MPH

[1]Department of Medicine, Emory University, Atlanta, GA, United States

[2]Emory Transplant Center, Atlanta, GA, United States

[3]Department of Medicine, Northwestern University, Chicago, IL, United States

[4]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

Corresponding Author:
Rachel E Patzer, PhD, MPH
Department of Medicine
Emory University
101 Woodruff Circle
Atlanta, GA,
United States
Phone: 1 404 727 6047
Email: rpatzer@emory.edu

Abstract

Background: Kidney and liver transplant recipients must manage a complex care regimen after kidney transplant. Although the use of Web-based patient portals is known to improve patient-provider communication and health outcomes in chronic disease populations by helping patients manage posttransplant care, disparities in access to and use of portals have been reported. Little is known about portal usage and disparities among kidney and liver transplant recipients.

Objective: The aim of this study was to examine patient racial/ethnic, socioeconomic, and clinical characteristics associated with portal usage among kidney and liver transplant recipients.

Methods: The study included all adult kidney and liver transplant recipients (n=710) at a large academic transplant center in the Southeastern United States between March 2014 and November 2016. Electronic medical record data were linked with Cerner portal usage data. Patient portal use was defined as any portal activity (vs no activity) recorded in the Cerner Web-based portal, including viewing of health records, lab results, medication lists, and the use of secure messaging. Multivariable log-binomial regression was used to determine the patient demographic, clinical, and socioeconomic characteristics associated with portal usage, stratified by organ.

Results: Among 710 transplant recipients (n=455 kidney, n=255 liver), 55.4% (252/455) of kidney recipients and 48.2% (123/255) of liver recipients used the patient portal. Black patients were less likely to use the portal versus white patients among both kidney (57% black vs 74% white) and liver (28% black vs 55% white) transplant recipients. In adjusted multivariable analyses, kidney transplant recipients were more likely to use the portal if they had higher education; among liver recipients, patients who were white versus black and had higher education were more likely to use the portal.

Conclusions: Despite studies showing that patient portals have the potential to benefit transplant recipients as a tool for health management, racial and socioeconomic disparities should be considered before widespread implementation. Transplant centers should include portal training and support to all patients to encourage use, given its potential to improve outcomes.

(J Med Internet Res 2019;21(4):e11864) doi:10.2196/11864

KEYWORDS

kidney transplantation; liver transplantation; patient portal; patient engagement; healthcare disparities

Introduction

Complexity of Transplant Care

Transplantation has improved health outcomes for patients with end-stage organ failure, doubling median survival rates compared with patients who remain on the waiting list for an organ (kidney 12.4 years vs 5.4 years; liver 11.6 years vs 3.1 years; heart 9.5 years vs 2.3 years; and lung 5.2 years vs 2.3 years) in the last 25 years [1-3]. Immunosuppressant therapy is essential to graft survival, and treatment regimens can be highly complex and a major burden for patients to manage [4]. Transplant recipients, for example, take an average of 11 medications per day and often experience frequent changes in their medication regimen [5,6]. In addition, because of the complex and multidisciplinary nature of their clinical management, transplant recipients are required to navigate complex systems of care by attending frequent clinical appointments, obtaining frequent laboratory and radiologic testing, and meeting with multiple medical specialists.

Potential Impact of Internet Technologies Upon Transplant Care

Modern internet technologies have been used to help manage other complex conditions, including diabetes, HIV, asthma, depression, heart disease, tuberculosis, and end-stage renal disease, leading to improvement in overall health and self-care management [7-13]. These new technologies have increased patient empowerment, decreased office visit rates and phone calls to physicians, and improved adherence to treatment [14,15]. However, little is known about the use of modern Web-based technologies among the transplant recipient patient population. Studies have indicated that owing to the high prescription burden, transplant recipients are interested in Web-based applications to assist in their complex medication management [16]. For example, Browning et al reported that 78% of adult kidney recipients had a positive attitude toward the use of mobile health smartphone apps for medication management [17].

Web-based patient portal systems can enhance convenience, communications, and fidelity in patient care. Despite previous research showing positive patient attitudes toward health technology, it may not be adopted equitably by all patients, and patient portal use is likely decreased in patients of minority race, lower educational level, and older age [18-21]. For example, non-Hispanic whites, younger adults, and those with high levels of education are more likely to use patient portals for diabetes management compared with racial and ethnic minorities, older adults, and those with lower education [18]. Limited access to technology may be a contributor to this gap. Other studies, however, suggest that the observed disparities are less related to access but that a lack of familiarity with data retrieval (eg, electronic health literacy) helps explain the lower portal usage among these disparate groups [22]. Among hospitalized patients with low socioeconomic status, inadequate communication between patients and providers during discharge and posthospital transition periods has also been cited as a potential barrier to medication access and adherence [23]. A patient portal has the potential to overcome barriers in patient-provider communication, but little is known about potential disparities in portal usage among kidney and liver transplant recipients.

The aim of this study was to examine the patient demographic, clinical, and socioeconomic characteristics associated with kidney and liver transplant recipients' usage of a patient portal among a diverse population transplanted at a large transplant center in the Southeastern United States. Results of this study may help transplant centers better understand which transplant patients use the portal as we seek to improve portal use as a means of improving patient-provider communications and health outcomes.

Methods

Study Design

In this cross-sectional study, we examined patient portal use among all adult patients (N=710) who received a single-organ kidney or liver transplant at a large transplant center in the Southeast between March 2014 and November 2016. Owing to significant clinical and programmatic differences between kidney and liver recipients, which may impact portal use, they were treated as 2 separate cohorts. The patient Web portal is part of the institutional electronic medical record (EMR) system and integrates clinical information into 1 accessible Web interface for patients. Patient information is summarized on the portal landing page, and tabs direct patients to specific functions such as health record overview, medication overview, appointment management, prescription renewal, and bill payment. Health records include laboratory results, which are updated within 36 hours, and radiology reports, which are updated within 7 days, after authorization and review by the ordering physician. The messaging function allows patients to communicate with providers and transplant nurse coordinators. Since the launch of the patient Web portal in March 2014, the patient portal was mentioned to kidney and liver candidates at evaluation for a transplant. At the time of evaluation for kidney candidates and before discharge after transplantation for both liver and kidney recipients, patients received instructions on how to access the portal and were informed that they would receive an email invitation prompting them to register for the portal.

Data Sources

Kidney and liver transplant recipients were identified using the transplant center's EMR data for demographic variables and linked with data from the national United Network for Organ Sharing (UNOS) database to collect additional demographic, clinical, and socioeconomic variables. All demographic, clinical, and socioeconomic variables were obtained from UNOS with the exception of marital status, which was obtained from the institution's EMR. If data were missing from UNOS, they were extracted from the EMR if available. Patient zip code was linked to US Census American Community Survey 5-year data from 2014 to collect zip code–level data on poverty. In addition, patient data were linked to patient portal activity, which was documented in Web server logs as part of the institution's EMR (Cerner Corp, North Kansas City, MO) and included each portal function performed with a timestamp.

Study Variables

The main outcome was patient usage of the Web portal, defined as whether a patient had any recorded activity based on Cerner server logs at any point during the study period, which includes 6 months before transplant and 2 years after transplant.

Demographic variables included age at transplant, gender, race/ethnicity (white, black, Hispanic, or other), marital status (single vs married), patient's highest level of education (categorized as grade and high school, some college, and college/graduate degree). Sociodemographic variables included insurance at the time of transplant (private vs public, which included Medicare and Medicaid), employment at wait-listing, percent of people living under the federal poverty line in the patient's zip code (categorized as <10%, 10 to 15%, or ≥15%), and patient distance from transplant center (distance between zip codes). Clinical characteristics included disease etiology, length of stay during time of transplant (calculated by subtracting transplant date from discharge date), functional status at the time of transplant measured by Karnofsky Performance Score (categorized as < vs ≥80%, which was defined as "Normal activity with effort: some symptoms of disease"). For kidney transplant recipients, we collected donor type (deceased or living). For liver transplant recipients, measures of liver disease severity at the time of transplant such as the laboratory model for end-stage liver disease (MELD) score and presence of encephalopathy were also collected. In the liver transplant cohort, we combined Hispanic race/ethnicity with the *other* race/ethnicity category for higher predictive power. All variables were obtained from UNOS.

In addition, exploratory outcomes were collected, including patient use of specific portal functions over the study period, which were measured using Cerner server logs documenting whether a patient used the portal to make appointments; to view clinical documents, current medications, current allergies, immunization records, visit history, radiology results, or lab results; or to engage in clinical messaging with providers. All time stamps were standardized in reference to the transplant date (t=0). Portal use was measured as a rate: number of clicks in the portal per 100 patients between March 2014 and November 2016.

Statistical Analyses

We compared patient characteristics of portal users versus nonusers by using chi-square tests for categorical variables and independent *t* tests or Wilcoxon Rank Sum tests for continuous variables. All variables that were significant in these tests were included in a multivariable model. Using backward selection, variables with a risk ratio of 1 in the multivariable model were excluded from the final model. For variables with >10% missing data, a missing category was included for analyses. In addition, descriptive statistics for frequency of portal use functions were calculated for kidney and liver patients.

Results

Characteristics of Patient Population

Patient characteristics of 710 kidney and liver transplant recipients included in the study population are shown in Tables 1 and 2, respectively, of which 64.1% (455/710) were kidney transplant recipients and 35.9% (255/710) were liver transplant recipients; results are stratified by portal usage. Portal usage was slightly higher in kidney transplant patients (55.4%, 252/455) compared with the liver cohort (48.2%, 123/255). The median age of kidney transplant recipients was 49, and the majority of patients were black (56.9%, 259/455) followed by white (33.9%, 154/455), Hispanic (4.4%, 20/455), and other race/ethnicity (4.8%, 22/455). More than half of the kidney recipients were male (55.2%, 251/455)), married (53.4%, 243/455) with the majority having public insurance (73.6%, 335/455), being unemployed at time of wait-listing (55.2%, 251/455), and receiving a deceased donor transplant (61.5%, 280/455). Among the liver transplant recipients, the median age was 53 and the majority of patients were male (59.6%, 152/255), white (71.8%, 183/255), followed by black (23.9%, 61/255), Hispanic (0.8%, 2/255), and other race/ethnicity (3.5%, 9/255). Most patients were married (65.5%, 167/255), had lower education (42.7%, 109/255 completed grade or high school only), and had private insurance (58.4%, 149/255).

Table 1. Characteristics of kidney transplant recipients between March 2014 and November 2016, by portal use.

Characteristics	Kidney			
	All (N=455)	User (N=252)	Nonuser (N=203)	*P* value
Age at transplant, median (IQR[a])	49.1 (13.0)	48.1 (19.1)	50.3 (24.8)	.08
Gender, n (%)				>.99
Male	251 (55.2)	139 (55.2)	112 (55.2)	—[b]
Female	204 (44.8)	113 (44.8)	91 (44.3)	—
Race/Ethnicity, n (%)				<.001
White	154 (33.9)	114 (45.2)	40 (19.7)	—
Black	259 (56.9)	120 (47.6)	139 (68.5)	—
Hispanic	20 (4.4)	5 (2.0)	15 (7.4)	—
Other	22 (4.8)	13 (5.2)	9 (4.4)	—
Marital Status, n (%)				0.03
Single	208 (45.7)	104 (41.3)	104 (51.2)	—
Married	243 (53.4)	147 (58.3)	96 (47.3)	—
Missing	4 (0.8)	1 (0.4)	3 (1.5)	—
Education new, n (%)				<.001
Grade and high school	147 (32.3)	51 (20.2)	97 (47.8)	—
Some college	152 (33.4)	86 (34.1)	66 (32.5)	—
College/graduate degree	147 (32.3)	110 (43.7)	37 (18.2)	—
Missing	8 (1.8)	5 (2.0)	3 (1.5)	—
Insurance category, n (%)				<.001
Public	335 (73.6)	160 (75.4)	175 (86.2)	—
Private	120 (26.4)	92 (36.5)	28 (13.8)	—
Employment at listing, n (%)				<.001
Yes	179 (39.3)	130 (51.6)	49 (24.1)	—
No	270 (59.3)	120 (47.6)	150 (73.9)	—
Missing	6 (1.3)	2 (0.8)	4 (2.0)	—
US Census poverty level by zip code, n (%)				<.001
>15% (poorest)	204 (44.8)	91 (36.1)	113 (55.7)	—
10-15%	100 (22.0)	60 (23.8)	40 (19.7)	—
<10% (wealthiest)	145 (31.9)	100 (39.7)	45 (22.2)	—
Missing	6 (1.3)	1 (0.4)	5 (2.5)	—
Distance (miles), median (IQR)	25 (14-68)	25 (17-57)	26 (13-79)	.81
Days on waiting list, median (IQR)	671 (258-1413)	608 (231-1,348)	798 (305-1,464)	.06
Length of stay, median (IQR)	4 (3-5)	4 (3-5)	4 (4-6)	<.001
Etiology (kidney), n (%)				<.001
Diabetes	86 (18.9)	43 (17.1)	43 (21.2)	—
Hypertension	156 (34.3)	61 (24.2)	95 (46.8)	—
Glomerulonephritis	117 (25.7)	83 (32.9)	34 (16.7)	—
Other	96 (21.1)	65 (25.8)	31 (15.3)	—

Characteristics	Kidney			
	All (N=455)	User (N=252)	Nonuser (N=203)	*P* value
Functional status group, n (%)				<.001
<80%	192 (42.2)	82 (32.5)	110 (54.2)	—
>80%	259 (56.9)	168 (66.6)	91 (44.8)	—
Missing	4 (0.9)	2 (0.8)	2 (1.0)	—
Donor type, n (%)				<.001
Deceased	280 (61.5)	125 (49.6)	155 (76.4)	—
Living	175 (38.5)	127 (50.4)	48 (23.6)	—

[a] IQR: interquartile range.

[b] Missing data or not applicable.

Patient Portal Use Among Kidney Transplant Recipients

About half (55.4%, 252/455) of kidney transplant recipients were portal users (Table 1). Portal usage was the same among males and females (55.2%, 139/252 and 112/203), but portal users were slightly younger than nonusers (median: 49 vs 51; *P*=.08). Only 46.3% (120/259) of black patients and 25% (5/20) of Hispanic patients used the portal compared with 74.0% (114/154) of white patients (*P*<.001). Married versus single patients (60.5%, 147/243 vs 50.0%, 104/208; *P*=.03), patients with college or graduate school education versus grade/high school education (74.8%, 110/147 vs 34.7%, 51/147; *P*<.001), patients with private versus public insurance (76.7%, 92/120 vs 47.8%, 160/335; *P*<.001), patients employed versus unemployed (72.6%, 130/179 vs 44.4%, 120/270; *P*<.001), and the highest zip code poverty–level group versus the lowest zip code poverty–level group (69.0%, 100/145 vs 44.6%, 91/204; *P*<.001) were more likely to use the portal. Portal users had shorter times on the kidney transplant waiting list compared with nonportal users (608 vs 798 days; *P*=.06). Patients able to perform normal activity upon discharge used the portal more often compared with patients who were not able to perform normal activity (64.9%, 168/259 vs 42.7%, 82/192; *P*<.001). Patients with a disease etiology of glomerulonephritis (70.9%, 83/117 used portal) or *other* cause (68%, 65/96 used portal) were more likely to be portal users compared with patients with diabetes (50%, 43/86) or hypertension (39%; *P*<.001). Patients with higher (≥80%) versus lower (<80%) functional status (64.9% (168/259) vs 42.7% (82/192); *P*<.001) and patients with a living versus deceased donor transplant (72.5%, 127/175 vs 44.6%, 125/280; *P*<.001) were also more likely to be portal users.

In multivariable adjusted log-binomial regression models, the following variables were included in the final model: race/ethnicity, education, employment, insurance, zip code poverty level, marital status, etiology, functional status, and donor type. Once adjusting for these covariates, only the education level remained significant (Table 3). Transplant recipients with a college or graduate degree were more likely to use the patient portal than patients who finished grade and high school only (adjusted risk ratio [aRR] 1.16; 95% CI 1.01-1.32).

Table 2. Characteristics of liver transplant recipients between March 2014 and November 2016, by organ type and portal use.

Characteristics	Liver			
	All (N=255)	User (N=123)	Nonuser (N=132)	*P* value
Age at transplant, median (IQR[a])	53.4 (11.5)	53.0 (11.7)	53.9 (11.3)	.64
Gender, n (%)				.27
Male	152 (59.6)	69 (56.1)	83 (62.9)	—[b]
Female	103 (40.4)	54 (43.9)	49 (37.1)	—
Race/Ethnicity, n (%)				.003
White	183 (71.8)	101 (82.1)	82 (62.1)	—
Black	61 (23.9)	17 (13.8)	44 (33.3)	—
Hispanic	2 (0.8)	1 (0.8)	1 (0.8)	—
Other	9 (3.5)	4 (3.3)	5 (3.8)	—
Marital status, n (%)				.91
Single	88 (34.5)	42 (34.1)	46 (34.8)	—
Married	167 (65.5)	81 (65.9)	86 (65.2)	—
Missing	0 (0.0)	0 (0.0)	0 (0.0)	—
Education new, n (%)				<.001
Grade and high school	109 (42.7)	40 (32.5)	69 (52.3)	—
Some college	49 (19.2)	30 (24.4)	19 (14.4)	—
College/graduate degree	51 (20.0)	36 (29.3)	15 (11.4)	—
Missing	46 (18.0)	17 (13.8)	29 (22.0)	—
Insurance category, n (%)				.12
Public	106 (41.6)	45 (36.6)	61 (46.2)	—
Private	149 (58.4)	78 (63.4)	71 (53.8)	—
Employment at listing, n (%)				.02
Yes	22 (8.6)	16 (13.0)	6 (4.5)	—
No	221(86.7)	103 (83.7)	118 (89.4)	—
Missing	12 (4.7)	4 (3.3)	8 (6.1)	—
US Census poverty level by zip code, n (%)				.69
>15% (poorest)	101 (39.6)	45 (36.6)	56 (42.4)	—
10-15%	67 (26.3)	34 (27.6)	33 (25.0)	—
<10% (wealthiest)	85 (33.3)	42 (34.1)	43 (32.6)	—
Missing	2 (0.8)	2 (1.6)	0 (0.0)	—
Distance (miles), median (IQR)	40 (21-87)	37 (21-78)	42 (20-99)	.73
Days on waiting list, median (IQR)	53 (7-128)	57 (13-137)	43 (5-122)	.11
Length of stay, median (IQR)	10 (8-18)	10 (7-18)	11 (8-19)	.21
MELD Score, mean (SD)	23.0 (9.4)	21.7 (9.3)	24.2 (9.4)	.03
Encephalopathy at transplant, n (%)				.37
None	178 (69.8)	91 (74.0)	87 (65.9)	—
2-Jan	67 (26.3)	28 (22.8)	39 (29.5)	—
4-Mar	10 (3.9)	4 (3.3)	6 (4.5)	—

Characteristics	Liver			
	All (N=255)	User (N=123)	Nonuser (N=132)	*P* value
Etiology (liver), n (%)				.08
Viral	91 (35.7)	37 (30.1)	54 (40.9)	—
Alcohol	48 (18.8)	29 (23.6)	19 (14.4)	—
Other	116 (45.5)	57 (46.3)	59 (44.7)	—
Functional status group, n (%)				.38
<80%	87 (34.1)	39 (31.7)	48 (36.3)	—
>80%	162 (63.5)	82 (66.6)	80 (60.6)	—
Missing	6 (2.4)	2 (1.6)	4 (3.1)	—

[a] IQR: interquartile range.

[b] Missing data or not applicable.

Table 3. Risk ratios for portal user versus nonusers among kidney transplant recipients between March 2014 and November 2016.

Characteristics	Multivariable model, adjusted risk ratio (95% CI)	*P* value
Race/Ethnicity		
White	Reference[a]	—[b]
Black	0.96 (0.84-1.09)	.49
Hispanic	0.83 (0.62-1.12)	.22
Other	0.92 (0.75-1.13)	.43
Married versus single	1.02 (0.93-1.12)	.63
Education level		.48
Grade school and high school	Reference	—
Some college	1.11 (0.98-1.26)	.09
College/graduate degree	1.16 (1.01-1.32)	.03
Employed versus unemployed	1.07 (0.96-1.19)	.23
Private versus public insurance	1.02 (0.91-1.15)	.69
US Census poverty level by zip code		
>15% (poorest)	Reference	—
10%-15 %	1.08 (0.95-1.22)	.23
<10% (wealthiest)	1.07 (0.95-1.20)	.24
Etiology		
Diabetes	1.04 (0.90-1.21)	.60
Hypertension	Reference	—
Glomerulonephritis	1.12 (0.98-1.27)	.09
Other	1.10 (0.95-1.27)	.19
>80% versus <80% function	1.06 (0.95-1.17)	.29
Living versus deceased donor	1.06 (0.94-1.19)	.34

[a] Reference: Comparison group to which all others within characteristic are compared with.

[b] Missing data or not applicable.

Patient Portal Use Among Liver Transplant Recipients

About half (48.2%, 123/255) of liver transplant recipients were portal users, with portal use being higher among females versus males (52.4%, 54/103 vs 45.4%, 69/152; *P*=.27) and white versus black patients (55.2%, 101/183 vs 27.9%, 17/61; *P*=.27; *P*=.003) (Table 2). Patients with college or graduate school education versus those who completed grade/high school education only (71%, 36/51 vs 36.7% 40/109; *P*<.001), patients with private versus public insurance (52.3%, 78/149 vs 42.5%, 45/106; *P*=.12), and patients employed versus those unemployed (73%, 16/22 vs 46.6% 103/221; *P*=.02) were more likely to use the portal. Patients who used the portal had longer times on the liver transplant waiting list compared with nonportal users (57 versus 43 days; *P*=.11) and lower MELD scores compared with nonportal users (21.7 versus 24.2; *P*=.03). Patients with an alcohol-related disease etiology (60%, 29/48) were more likely to use the portal compared with patients with a viral cause (41%, 37/91) or cause categorized as other (49.1%, 57/116; *P*=.08). Age, marital status, zip code poverty level, and distance were similar among portal and nonportal users among this cohort.

In multivariable adjusted binomial regression models, gender, race/ethnicity, marital status, education, employment, insurance, etiology, and MELD score were included in the final model (Table 4). In the final adjusted model, only race/ethnicity and education level remained significant. Black patients were significantly less likely than white patients to utilize the portal (adjusted RR 0.65; 95% CI 0.46-0.92). Transplant recipients with some college education or with a college or graduate degree were more likely to use the portal compared with those with a grade or high school education (some college versus grade/high school adjusted RR 1.36; 95% CI 1.01-1.84; college/graduate degree versus grade/high school adjusted RR 1.46; 95% CI 1.08-1.98; Table 4).

Table 4. Risk ratios for portal user versus nonusers among liver transplant recipients between March 2014 and November 2016.

Comparison groups	Multivariable model, adjusted risk ratio (95% CI)	*P* value
Female versus male	1.03 (0.83-1.29)	.76
Race/Ethnicity		
White	Reference[a]	—[b]
Black versus white	0.65 (0.46-0.92)	.01
Other versus white	0.84 (0.53-1.32)	.44
Education level		
Grade and high school	Reference	—
Some college versus grade and high school	1.36 (1.01-1.84)	.04
College/graduate degree versus grade and high school	1.46 (1.08-1.98)	.01
Missing versus grade and high school	0.94 (0.64-1.38)	.76
Employed versus unemployed	1.11 (0.81-1.53)	.50
Private versus public insurance	1.03 (0.80-1.33)	.83
Etiology		
Viral	0.84 (0.64-1.11)	.22
Alcohol	1.14 (0.86-1.50)	.36
Other	Reference	—
Model for end-stage liver disease score	0.99 (0.97-1.00)	.08

[a] Reference: Comparison group to which all others within characteristic are compared with.

[b] Missing data or not applicable.

Portal Function Usage

In exploratory analyses among portal users, we found that the frequency of overall portal activity among both kidney and liver transplant recipients was similar; however, before transplantation, liver transplant recipients used the portal more frequently (Figure 1). Among both kidney and liver cohorts, there was an increase in recorded portal activity from 6 months (–150) before the time of transplant (t=0; approximately 180 clicks/100 kidney patients and approximately 140 clicks/100 liver patients), which gradually started declining 45 days after transplant. As portal activity gradually increased before transplantation in liver recipients, a more dramatic increase in portal clicks posttransplant was recorded in kidney recipients with more sustained use in the first 400 days of transplant (Figure 1).

Across the entire study period, portal functions were used at different rates among both kidney and liver transplant recipients (Table 5).

Viewing lab results was the most frequent function used over the study period for kidney and liver patients (43.9% and 37.0%, respectively), followed by viewing immunizations (18.1% and 27.9%, respectively), allergies (18.6% and 20.2%, respectively), and messaging (12.0% and 5.2%, respectively; Figure 2). Portal activity for the different viewing functions was similar among kidney and liver transplant recipients, although kidney recipients tended to have higher activity viewing messages compared with liver patients, and liver patients had higher activity viewing immunizations compared with kidney patients.

Figure 1. Overall portal activity among kidney (n=252) and liver (n=123) transplant recipients who registered for the patient portal between March 2014 and November 2016, 6 months before transplant to 2 years after transplant. Across the entire study period, portal functions were used at different rates among both.

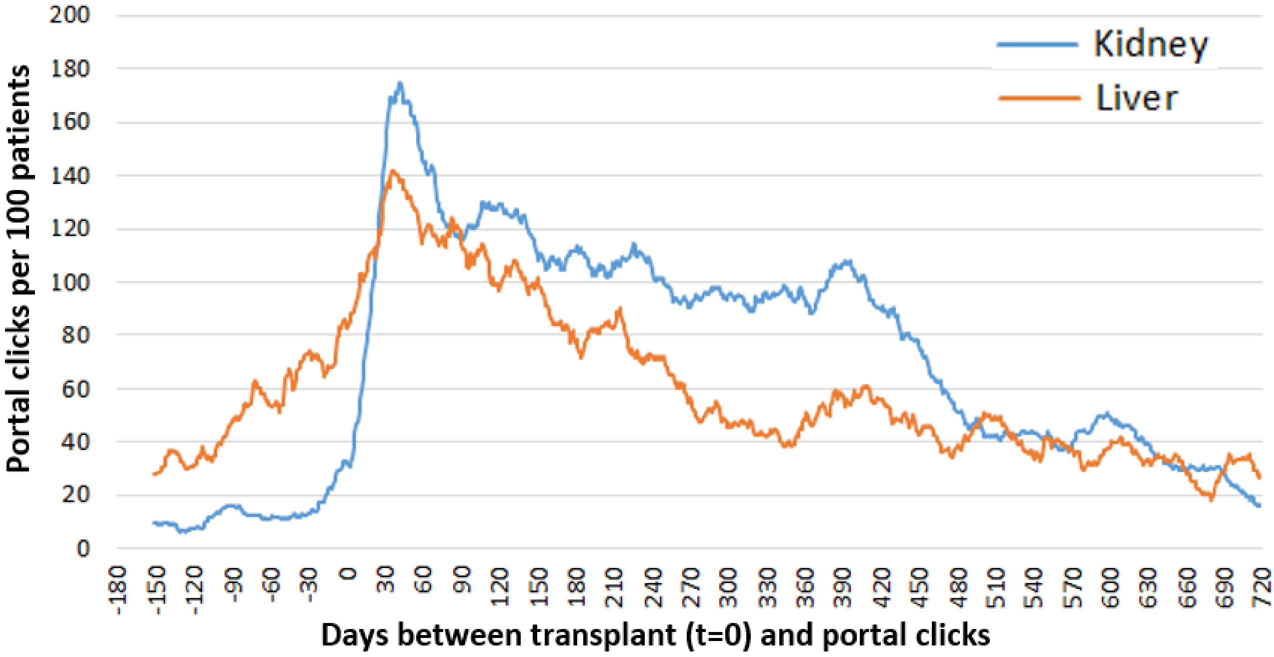
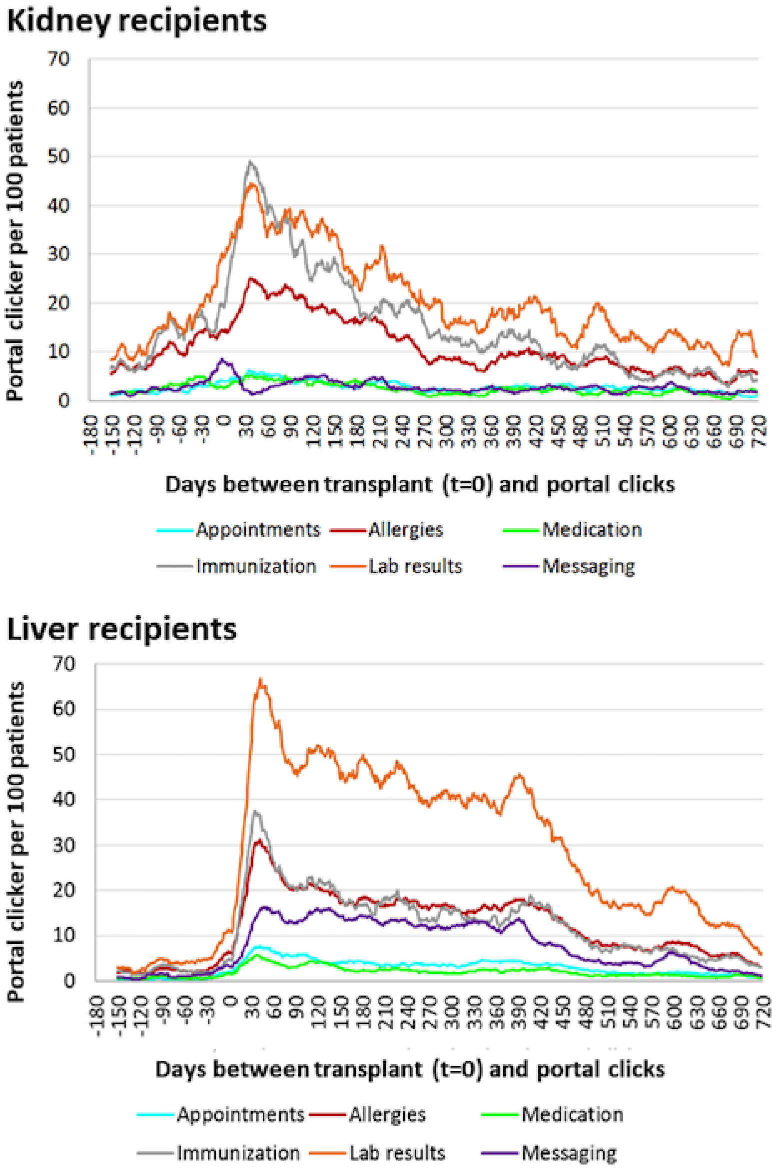


Table 5. Portal functions used among kidney (n=252) and liver (n=123) transplant recipients who registered for the patient portal between March 2014 and November 2016, 6 months before transplant to 2 years after transplant. Percentages do not sum up to 100% due to rounding errors.

Portal functions	Kidney, n (%)	Liver, n (%)
Appointments	11 (4.4)	6 (5.1)
Allergies	47 (18.6)	25 (20.2)
Medication list	7 (2.9)	6 (4.5)
Immunizations	46 (18.1)	34 (27.9)
Lab results	111 (43.9)	46 (37.0)
Messages	30 (12.0)	6 (5.2)

Figure 2. Portal function activity viewed among kidney (n=252) and liver (n=123) recipients who registered for the patient portal between March 2014 and November 2016, 6 months before transplant to 2 years after transplant.



Discussion

In this single center cross-sectional study among a diverse population of kidney and liver transplant recipients at a major transplant center in the Southeast, Web-based patient portal usage was relatively low. Only about half of the patients used a Web-based patient portal at least one time after receiving a transplant, though previous literature has shown that this patient population is generally receptive to technological tools to help manage their complex medication regimes [17]. This study is the first to examine both kidney and liver transplant patients' interaction with a Web-based patient portal, specifically with regard to the influence of socioeconomic characteristics on portal usage. In addition, this study uses novel data sources, including Web server logs to verify actual portal usage.

We identified significant racial and socioeconomic disparities in Web-based patient portal usage. Among kidney transplant recipients, patients with lower education levels were significantly less likely to use the portal. Among liver transplant recipients, black versus white patients and patients with lower education levels were less likely to use the portal. These results suggest that as patient portals have the potential for improvement in self-care management, strategies need to be developed to improve usage among minority patients and patients with lower education levels.

Certain findings in the multivariable analysis are not as clearly related to socioeconomic factors, and as there is sparsity in the literature of portal usage in transplant recipients, we speculate upon possible explanations. Kidney recipients with hypertension and diabetes as the cause of kidney disease were less likely to use the portal compared with those with other etiologies. In addition, among kidney recipients, living donor type and shorter length of stay during transplant were associated with increased portal usage. Kidney recipients receiving living donor kidney grafts may be less sick at transplant compared with deceased donor recipients who have spent more time on dialysis and thus less likely to use the portal. Similarly, patients with a longer length of hospital stay may be more complicated. Thus, degree of illness and complicated medical comorbidities may be markers for decreased portal use. As potential kidney recipients experience difficulty directly asking others to donate, kidney recipients with available living donors may have a more complete and comfortable social support network [24], which may also result in greater use of the patient portal by multiple family members. Liver recipients with higher laboratory MELD scores at transplant were also less likely to use the portal, again suggesting that the degree of illness at the time of transplant is associated with portal usage. Alternatively, high MELD patients are often acutely ill and newer to the system, limiting their familiarity with the portal. These findings suggest that patients who are more ill at the time of transplant may need to have transplant center staff reintroduce Web portals and their potential usefulness during subsequent appointments after transplant. In addition, patients who are more ill may benefit from more caregiver involvement.

Other studies have identified disparities in Web-based patient portal usage among patients with other chronic conditions [18,20,25-28]. For example, Goel et al reported a large racial/ethnic disparity in enrollment of their patient portal (Northwestern Medical Faculty Foundation) among a general internal medicine clinic population [20]. Roblin et al examined portal usage among adults with diabetes, adults with elevated lipids but no history of advanced coronary artery disease, and *low risk* adults in the Kaiser Permanente Georgia population, and found a similar distribution of racial and socioeconomic characteristics in portal usage, with socioeconomic characteristics not accounting for the disparities in usage by race/ethnicity [25]. In a separate study of adult diabetic patients, black and Hispanic patients had the highest odds of never logging on to the patient portal compared with non-Hispanic white patients [18]. Similar to our study, this study and others regarding patients with other conditions have shown that patients with lower socioeconomic status are less likely to access electronic health information [26].

Overall, the portal was used most within the first 30 days of transplant, but portal use declined after that time. Viewing lab results in the portal was the most used function by both kidney and liver patients. Portal activity was higher among patients awaiting liver transplant compared with those awaiting kidney transplant; however, after transplant, kidney recipients' use increased above liver recipients' use and was steady through, approximately, the next 500 days posttransplant. The reasons and implications of this pattern are unclear. Variations in clinic protocols between liver and kidney candidates before transplant may explain increased use in the liver recipients because pretransplant clinic interaction tends to be more rigorous for liver patients whereas kidney patients are often primarily managed by their preexisting nephrologist before transplant. Conversely, introduction to the patient portal included an instructional packet for kidney candidates but did not do so for liver candidates. However, this study does not explain the greater portal use by kidney patients after transplant. Use of the medication list function was low for both groups of recipients. Although the low usage may be related to center-specific protocols using alternative methods for medication reconciliation and communication, supporting the patient use of this function could translate to more accurate chart records of medications and could be the focus of future portal-based interventional studies.

Although internet access has been shown to be a contributing factor to racial disparities in portal usage, these disparities have decreased in the past decade; the proportion of whites, blacks, and Hispanics using the internet is 88%, 85%, and 88%, respectively [29]. Socioeconomic disparities are associated with health literacy challenges, so limited health literacy may both explain disparities in portal use but also help to identify individuals who could benefit most from portal-based initiatives [30]. This study suggests that when designing and implementing portals for transplant candidates and recipients, centers should consider socioeconomic factors, such as lower education levels, that may limit exposure of the portal to its target user and increase the need for targeted outreach and education about use and benefit of the portal.

There are several limitations to our study. Regional racial and socioeconomic composition is a limiting factor to the generalizability of this study. Among kidney transplant recipients, the proportion of African-Americans was higher compared with the proportion nationally reported in the OPTN/SRTR 2015 Annual Kidney Data Report (57% study vs 28% national) [3]. However, similar to the national data, kidney recipients were mostly publicly insured (74% study vs 67% national). Among liver transplant recipients, compared with the national data, the percentage of whites is similar (71% national vs 72% study), but the percentage of blacks is more than double the national rate (10% national vs 24% study), and most patients were privately insured (54% national vs 58% study) [2]. Other limitations of our study include lack of patient-reported data on the acceptability of portal use, lack of prospective follow-up to ascertain associations between portal use and outcomes, and lack of granular data about other potential socioeconomic variables associated with portal use. Findings from this study support the need for prospective and interventional studies using the patient portal in transplant populations.

On the basis of findings from this study and the other studies previously described, disparities in use of Web-based technologies seem to be universal among various health conditions, software systems, and US regions, implying a developmental flaw. Patient portals, as tools for health management, have the potential to benefit transplant recipients. However, racial and socioeconomic disparities should be considered during the development and implementation of any digital tool, including the development of innovative strategies to increase exposure and acceptance to a patient portal.

Acknowledgments

This work was supported by National Institutes of Health (NIH) grant number R21NR014544. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. Rana A, Gruessner A, Agopian VG, Khalpey Z, Riaz IB, Kaplan B, et al. Survival benefit of solid-organ transplant in the United States. JAMA Surg 2015 Mar 1;150(3):252-259. [doi: 10.1001/jamasurg.2014.2038] [Medline: 25629390]
2. Kim W, Lake J, Smith J, Skeans M, Schladt D, Edwards E, et al. OPTN/SRTR 2015 Annual Data Report: liver. Am J Transplant 2017 Dec;17(Suppl 1):174-251 [FREE Full text] [doi: 10.1111/ajt.14126] [Medline: 28052604]
3. Hart A, Smith J, Skeans M, Gustafson S, Wilk AR, Robinson A, et al. OPTN/SRTR 2016 Annual Data Report: kidney. Am J Transplant 2018 Jan;18(Suppl 1):18-113 [FREE Full text] [doi: 10.1111/ajt.14557] [Medline: 29292608]
4. Promraj R, Dumronggittigule W, Sirivatanauksorn Y, Ruenrom A, Tovikkai C, Limsrichamrern S, et al. Immunosuppressive medication adherence in liver transplant recipients. Transplant Proc 2016 May;48(4):1198-1201. [doi: 10.1016/j.transproceed.2015.12.097] [Medline: 27320586]
5. Serper M, Patzer RE, Reese PP, Przytula K, Koval R, Ladner DP, et al. Medication misuse, nonadherence, and clinical outcomes among liver transplant recipients. Liver Transpl 2015 Jan;21(1):22-28 [FREE Full text] [doi: 10.1002/lt.24023] [Medline: 25312406]
6. Patzer RE, Serper M, Reese PP, Przytula K, Koval R, Ladner DP, et al. Medication understanding, non-adherence, and clinical outcomes among adult kidney transplant recipients. Clin Transplant 2016 Dec;30(10):1294-1305 [FREE Full text] [doi: 10.1111/ctr.12821] [Medline: 27447351]
7. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
8. Lima IC, Galvão MT, Alexandre HD, Lima FE, Araújo TL. Information and communication technologies for adherence to antiretroviral treatment in adults with HIV/AIDS. Int J Med Inform 2016 Aug;92:54-61. [doi: 10.1016/j.ijmedinf.2016.04.013] [Medline: 27318071]
9. Hui C, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 1;24(3):619-632. [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
10. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013;8(7):e62873 [FREE Full text] [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]
11. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
12. Molton JS, Pang Y, Wang Z, Qiu B, Wu P, Rahman-Shepherd A, et al. Prospective single-arm interventional pilot study to assess a smartphone-based system for measuring and supporting adherence to medication. BMJ Open 2016 Dec 20;6(12):e014194 [FREE Full text] [doi: 10.1136/bmjopen-2016-014194] [Medline: 27998903]
13. Patzer R, Basu M, Larsen C, Pastan S, Mohan S, Patzer M, et al. iChoose kidney: a clinical decision aid for kidney transplantation vs. dialysis treatment. Transplantation 2016 Mar;100(3):630-639 [FREE Full text] [doi: 10.1097/TP.0000000000001019] [Medline: 26714121]
14. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
15. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
16. Israni A, Dean C, Kasel B, Berndt L, Wildebush W, Wang CJ. Why do patients forget to take immunosuppression medications and miss appointments: can a mobile phone app help? JMIR Public Health Surveill 2016;2(1):e15 [FREE Full text] [doi: 10.2196/publichealth.5285] [Medline: 27227150]
17. Browning R, McGillicuddy J, Treiber F, Taber D. Kidney transplant recipients' attitudes about using mobile health technology for managing and monitoring medication therapy. J Am Pharm Assoc (2003) 2016;56(4):450-4.e1 [FREE Full text] [doi: 10.1016/j.japh.2016.03.017] [Medline: 27450140]

18. Sarkar U, Karter A, Liu J, Adler N, 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 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
19. 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]
20. 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]
21. Goel M, Brown T, Williams A, Cooper A, Hasnain-Wynia R, Baker D. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
22. Connolly K, Crosby M. Examining e-Health literacy and the digital divide in an underserved population in Hawai'i. Hawaii J Med Public Health 2014 Feb;73(2):44-48 [FREE Full text] [Medline: 24567867]
23. Cope R, Jonkman L, Quach K, Ahlborg J, Connor S. Transitions of care: medication-related barriers identified by low socioeconomic patients of a federally qualified health center following hospital discharge. Res Social Adm Pharm 2018 Dec;14(1):26-30. [doi: 10.1016/j.sapharm.2016.12.007] [Medline: 28096051]
24. Rodrigue JR, Cornell DL, Kaplan B, Howard RJ. Patients' willingness to talk to others about living kidney donation. Prog Transplant 2008 Mar;18(1):25-31. [doi: 10.1177/152692480801800107] [Medline: 18429579]
25. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
26. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Dec;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
27. Gordon N, Hornbrook M. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 4;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
28. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
29. Pew Research Center. 2017. Internet/Broadband Fact Sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2019-02-25] [WebCite Cache ID 76SXnFbka]
30. Mark KE, Ying J, Christine P. National Center for Education Statistics. Washington DC; 2006. The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2019-02-26] [WebCite Cache ID 76TZfAL1u]

Abbreviations

EMR: electronic medical record
MELD: model for end-stage liver disease
NIH: National Institutes of Health
RR: risk ratio
UNOS: United Network for Organ Sharing

Edited by G Eysenbach; submitted 07.08.18; peer-reviewed by ER Pfaff, B Nørgaard; comments to author 08.11.18; revised version received 02.01.19; accepted 06.01.19; published 22.04.19.

Please cite as:
Wedd J, Basu M, Curtis LM, Smith K, Lo DJ, Serper M, Wolf MS, Parker R, Patzer RE
Racial, Ethnic, and Socioeconomic Disparities in Web-Based Patient Portal Usage Among Kidney and Liver Transplant Recipients: Cross-Sectional Study
J Med Internet Res 2019;21(4):e11864
URL: https://www.jmir.org/2019/4/e11864/
doi:10.2196/11864
PMID:31008707

Original Paper

Improving Moderator Responsiveness in Online Peer Support Through Automated Triage

David N Milne[1,2], PhD; Kathryn L McCabe[2,3,4], PhD; Rafael A Calvo[2,5], PhD

[1]School of Information, Systems and Modelling, Faculty of Engineering and Information Technology, University of Technology, Sydney, Sydney, Australia

[2]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia

[3]Department of Psychiatry and Behavioral Sciences, University of California (Davis), Davis, CA, United States

[4]Medical Investigation of Neurodevelopmental Disorders Institute, University of California (Davis), Davis, CA, United States

[5]Dyson School of Design Engineering, Imperial College, London, United Kingdom

Corresponding Author:
David N Milne, PhD
School of Information, Systems and Modelling
Faculty of Engineering and Information Technology
University of Technology, Sydney
PO Box 123
Sydney, NSW 2007
Australia
Phone: 61 (02) 9514 3171
Email: david.milne@uts.edu.au

Abstract

Background: Online peer support forums require oversight to ensure they remain safe and therapeutic. As online communities grow, they place a greater burden on their human moderators, which increases the likelihood that people at risk may be overlooked. This study evaluated the potential for machine learning to assist online peer support by directing moderators’ attention where it is most needed.

Objective: This study aimed to evaluate the accuracy of an automated triage system and the extent to which it influences moderator behavior.

Methods: A machine learning classifier was trained to prioritize forum messages as green, amber, red, or crisis depending on how urgently they require attention from a moderator. This was then launched as a set of widgets injected into a popular online peer support forum hosted by ReachOut.com, an Australian Web-based youth mental health service that aims to intervene early in the onset of mental health problems in young people. The accuracy of the system was evaluated using a holdout test set of manually prioritized messages. The impact on moderator behavior was measured as response ratio and response latency, that is, the proportion of messages that receive at least one reply from a moderator and how long it took for these replies to be made. These measures were compared across 3 periods: before launch, after an informal launch, and after a formal launch accompanied by training.

Results: The algorithm achieved 84% f-measure in identifying content that required a moderator response. Between prelaunch and post-training periods, response ratios increased by 0.9, 4.4, and 10.5 percentage points for messages labelled as crisis, red, and green, respectively, but decreased by 5.0 percentage points for amber messages. Logistic regression indicated that the triage system was a significant contributor to response ratios for green, amber, and red messages, but not for crisis messages. Response latency was significantly reduced (*P*<.001), between the same periods, by factors of 80%, 80%, 77%, and 12% for crisis, red, amber, and green messages, respectively. Regression analysis indicated that the triage system made a significant and unique contribution to reducing the time taken to respond to green, amber, and red messages, but not to crisis messages, after accounting for moderator and community activity.

Conclusions: The triage system was generally accurate, and moderators were largely in agreement with how messages were prioritized. It had a modest effect on response ratios, primarily because moderators were already more likely to respond to high priority content before the introduction of triage. However, it significantly and substantially reduced the time taken for moderators to respond to prioritized content. Further evaluations are needed to assess the impact of mistakes made by the triage algorithm and how changes to moderator responsiveness impact the well-being of forum members.

(J Med Internet Res 2019;21(4):e11410) doi:10.2196/11410

KEYWORDS

social support; triage; classification; natural language processing

Introduction

When facing tough times, often the best people to turn to are those who have been through similar challenges and who can provide empathy and support that is grounded in personal experience [1]. Asynchronous text-based forums are a common method for facilitating such peer support online and have been shown to reduce symptoms of distress [2] and improve one's sense of empowerment [3]. Their online nature allows individuals to access help at any time, from any location, with minimal cost and effort [4]. They can often be accessed anonymously to mitigate the fear of stigma that can be a barrier to help-seeking, particularly among the young [5].

Although online communities have much to offer, there exist potential pitfalls. For instance, they often lack the involvement of mental health professionals; community members' interactions may be influenced negatively by an individual's current mental health status [6]; social difficulties may be exacerbated online because of missing social cues [7] or illness-related disinhibition or disorganization [8]. Furthermore, without appropriate oversight, risky and unsafe behaviors can emerge unchecked, such as the normalization of self-harm, [9] and there is some evidence that online peer support can be misused as a method of avoidance [10].

The involvement of mental health professionals and paraprofessionals (ie, trained volunteers) likely improves the safety and therapeutic value of online peer support [11-13]. Users of these online communities appear to be amenable to oversight; for example, Kummervold et al [14] obtained almost unanimous feedback that mental health professionals should actively participate in online conversations and/or provide passive safety monitoring. Outside of peer support, increased moderation of online communities in general has been shown to improve intention to participate [15] and the quality of contributions [16].

The key barriers to greater oversight of online peer support are cost and scalability. As these communities grow, they place a greater burden on their human caretakers, which increases the likelihood that people in need may be overlooked. To address these barriers, we described an automated triage system that aims to guide human moderators to the people whose messages most urgently require their attention. We evaluated the accuracy with which this system identifies urgent content and the extent to which it influences moderator behavior. The evaluation of accuracy was conducted using a dataset of manually prioritized forum messages. The evaluation of behavior change was conducted as a quasi-experimental time series analysis that tracked moderator behavior for several years before and one year following the introduction of the triage system.

These evaluations were made within the context of ReachOut.com, an Australian Web-based youth mental health service that aims to intervene early in the onset of mental health problems in young people. ReachOut.com is well known and popular among its target population. Approximately 1 in 3 young people in Australia are aware of the site [17], and in 2015 the website received about 1.58 million Australian visitors [18]. In a survey conducted in 2013, approximately 77% of visitors reported experiencing high or very high levels of psychological distress, which indicates that the site is reaching people in need [19]. The site hosts a lively peer support forum for those aged 14 to 25 years to seek help and share their experiences. This community is maintained by staff employed by the organization and young volunteers who are recruited and trained for the role. Staff and volunteers—collectively referred to as the Mod Squad—monitor posts and respond as needed with encouragement, compassion, and referrals to relevant resources. This study investigated how to ensure the moderation provided by these professionals and volunteers remains scalable.

Methods

This section describes the triage system that was deployed and the method by which it was evaluated.

The Triage System

The triage system automatically prioritizes each new forum message as belonging to one of the following 4 categories:

- *Green* indicates a message can be safely left for the community to address, without requiring intervention from a moderator. Most forum messages are expected to fall into this category.
- *Amber* indicates a message that is important, but not urgent. It is appropriate for the moderator to wait and see if the community will respond to it. If no response is forthcoming, then a moderator should intervene.
- *Red* indicates a message that should be responded to as soon as possible, likely because the author is in distress or the message content may be triggering to others.
- *Crisis* indicates the author or someone they know is at risk of harm. A moderator should respond as soon as possible and enact ReachOut.com's existing escalation protocol if appropriate.

The triage system is embedded into the forum as a sidebar that provides a *to-do list* of crisis, red, and amber messages that have not received a response from a moderator. Additional widgets are also embedded below each forum message to display the priority assigned to it and whether it requires attention. Further details (including a screenshot) can be found in Multimedia Appendix 1.

The underlying algorithm that assigns these priorities relies on supervised machine learning, meaning that it learns automatically from examples of manually prioritized forum messages. The advantage of this approach—as opposed to manually specified rules—is that it can easily be adapted to new prioritization schemes or new online communities simply by feeding it new examples of prioritized messages. The algorithm can also easily be maintained by learning from any corrections it is given by moderators. Further details of the algorithm—and the features it relies on—can be found in Multimedia Appendix 2.

Evaluation of Accuracy

The evaluation of accuracy was conducted using a test set of manually prioritized messages that were withheld from any training or tuning of the algorithm. Both training and testing data were sourced from the Computational Linguistics and Clinical Psychology (CLPsych) 2016 shared task [20], which provided 1227 messages (947 for training and 280 for testing) that were extracted from ReachOut.com forums and manually labelled (ie, given one of the 4 priority labels described above) by 3 independent annotators. A reliability analysis indicated that these annotators reached a Fleiss kappa of 0.706 and pairwise Cohen kappa scores ranging between 0.674 and 0.761, indicating that though the task is somewhat subjective, a reasonable level of inter-rater agreement was achieved.

The primary evaluation metric used was f-measure, or the harmonic mean of recall (ie, sensitivity) and precision (ie, positive predictive value). As this is a multiclass classification problem where rare classes (eg, red and crisis) are of greater interest, scores were macro-averaged across the classes, after excluding the majority class (ie, green). We also reported the algorithm's performance in separating out content that is flagged for the moderators' attention (ie, amber, red, or crisis) from content that can safely be left for the community to address (ie, green), and in separating urgent content (ie, red and crisis) from content that can safely wait (ie, green and amber). In both cases we reported f-measure of the minority class (ie, flagged or urgent). Interested readers are directed to Milne et al's study [20] for more detailed information about the dataset and evaluation metrics.

Evaluation of the Impact on Moderator Response Behavior

A quasi-experimental time series analysis was used to compare moderator behavior before and after the introduction of the triage system. The primary measures of interest were response ratio and response latency: that is, the proportion of messages that received at least 1 reply from a moderator and the time elapsed between a message being created and receiving its first reply from a moderator. The analysis considers only replies made by moderators to messages authored by peers (ie, ordinary community members). All messages made by other types of forum members (eg, trainee moderators and other affiliates) are ignored, as are messages made by moderators unless in response to a message from a peer.

These measures were compared across 3 distinct periods: *prelaunch*, *postlaunch*, and *post-training*. The prelaunch period captured moderator behavior before the introduction of the triage system—from July 19, 2012, to August 5, 2016 (ie, 1478 days). The postlaunch period captured the interim in which the triage system was available, but not accompanied by any guidance about how it should be used or how often it should be consulted. It extended from August 5, 2016, to November 27, 2016 (ie, 114 days). The post-training period captured moderator behavior when the triage system was fully integrated into their workflow, having been launched with a detailed training session for all moderators. It extended from November 27, 2016, to October 16, 2017 (ie, 323 days).

During these periods, all priorities (ie, green, amber, red, or crisis labels) were assigned automatically. During the postlaunch and post-training periods, they were assigned immediately as each message was created and immediately revealed to users of the triage system. For the prelaunch period, they were assigned retroactively, simply by running the triage algorithm over the previously collected messages. These priority labels were not revealed to the moderators until after the launch of the triage system, and thus were unable to influence their behavior during the prelaunch period.

The volume of content that required moderation varied over time and could potentially have a strong effect on moderator response behavior. To account for this variability, we constructed histograms for each message that captured activity levels during 5 hour–long periods starting 2 hours before the message was created, the hour that the message was posted, and ending 3 hours afterward. These histograms record the number of messages created and the number of unique authors separately for peers and moderators. Together these 4 histograms measure the load placed on moderators (ie, the level of activity of the forum) and the number of active moderators available to share that load.

For further details about how moderators were identified, how replies were tracked for messages, and how the activity histograms were constructed, please refer to Multimedia Appendix 3.

Statistical Analysis

The triage algorithm was trained and evaluated using the scikit-learn Python Library, and the evaluation of moderator behavior was conducted using SPSS version 25 (IBM) statistical software for Mac. To assess whether the presence of the triage system and moderators' training with it (ie, period) were significant factors for moderator response ratio, direct logistic regression was performed separately for each priority level (green, amber, red, and crisis), while also accounting for moderator and peer activity at and around the time the message was posted. The impact of the triage system and moderator's training on moderator response latency was similarly evaluated using separate linear regression models for each priority level. Kolmogorov-Smirnov values and visual inspection indicated that response latency had a non-normal distribution, which was corrected using log transformation. Follow-up pairwise group differences were examined where significant overall tests were reported. Finally, to reduce the likelihood of type-I error, adjusted alpha levels were applied to account for multiple comparisons (P=.004).

Ethical Approval

This research was approved by the Human Research Ethics Committee at the University of Sydney as protocol 2016/064.

Results

Algorithm Accuracy

Table 1 shows the performance of the triage algorithm using the dataset and official metrics provided by the CLPsych 2016 shared task [20]. The system outperformed previous task participants in macroaveraged f-measure, ranked third in separating flagged (ie, crisis, red, or amber) messages from unflagged (ie, green) messages, and fourth in separating urgent messages (ie, crisis or red) from nonurgent (ie, amber or green) ones.

Volume and Severity of Messages

Table 2 provides a summary of the number of messages posted by peers within each evaluation period and how the algorithm automatically prioritized them. As expected, within each period (ie, each column), the majority of messages were green, and there were progressively fewer messages as priority increased.

Figure 1 provides a more detailed timeline, with a stacked histogram of the weekly volume of messages posted by peers, split into each of the 4 priorities. The number of messages posted each week fluctuated strongly, but there is a general upward trend indicating that the online community was becoming busier. At its peak, the forum received 1063 messages from peers during the week starting on July 2, 2017.

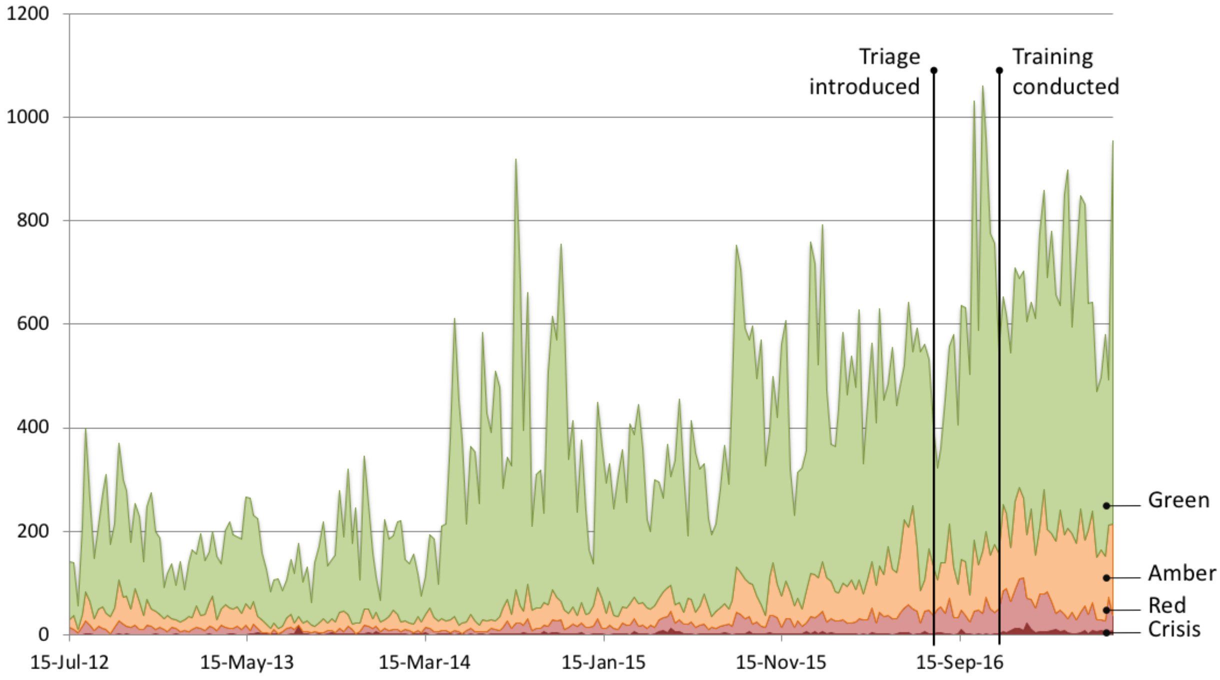
Table 1. Performance of the triage classifier and top participants of Computational Linguistics and Clinical Psychology (CLPsych) 2016.

Classification Algorithm	Macroaveraged *F*1	Flagged *F*1	Urgent *F*1
Kim et al [21]	0.42	0.85	0.62
Malmasi et al [22]	0.42	0.87	0.64
Brew [23]	0.42	0.78	0.69
Deployed to ReachOut.com	0.65	0.84	0.60

Table 2. Breakdown of how messages were prioritized (totals may be greater than 100% because of rounding).

Priority	Prelaunch, n (%)	Postlaunch, n (%)	Post-training, n (%)
Green	50,662 (80.1)	7808 (76.7)	22,008 (70)
Amber	8835 (14)	1634 (16)	6955 (22.1)
Red	3390 (5.4)	704 (6.9)	2170 (6.9)
Crisis	382 (0.6)	37 (0.4)	320 (1)

Figure 1. Weekly counts of prioritized messages from ordinary forum members.



Proportion of Messages That Received a Moderator Response

Table 3 provides a summary of how often messages from peers received at least 1 reply from a moderator. The data are organized by triage-assigned message priority, and the columns indicate whether the triage system was deployed at the time and whether training had been conducted.

Within each period (ie, column), there is a consistent progression in which the likelihood of a message receiving a reply is proportional to the priority assigned to it. The response ratios for green and red messages decrease after the introduction of triage but recover strongly after training. Response ratios for amber messages also decrease after the introduction of triage but do not recover after training. Response ratios for crisis messages change very little across the periods.

Logistic regression was performed to investigate whether these differences in response ratio were significant and to what extent they were because of the introduction of the triage system. Separate logistic regression models were completed for each priority label to predict the likelihood of a message receiving a moderator response. Each model contained 21 independent variables, including whether the triage system had been launched at the time of the message, whether training had been conducted, and the level of activity of moderators and community members at and around the time of the message (ie, the histogram data described previously).

For *green* messages, the model was significant (χ^2_{22}=11,411; N=80,478, $P<.001$), indicating that it was able to distinguish between messages that did and did not receive a moderator response. The model as a whole (ie, including all variables) explained between 13.3% (Cox and Snell R^2) and 18.2% (Nagelkerke R^2) of the variance in moderator response and correctly classified 70.2% of messages. In the final model (ie, retaining only statistically significant variables), the strongest predictor of moderator response was whether training for the triage system had been conducted, which recorded an odds ratio (OR) of 1.77. This was followed by the number of moderators online at the time of the community member message (OR 1.1). This indicated that green messages were more likely to receive a response after the triage system was introduced *and* moderators had been trained to use it after controlling for all other factors in the model.

For *amber* messages, the model was significant (χ^2_{22}=1207.2; N=17,424, $P<.001$). It explained between 6.7% (Cox and Snell R^2) and 9.4% (Nagelkerke R^2) of the variance in moderator response and correctly classified 70.3% of messages. The strongest predictor was whether triage training had been conducted recording an OR of 1.8. Again, this was followed by the number of moderators online at the time of the community member message (OR 1.3). This indicates that amber community messages were 1.8 times more likely to receive a moderator response after triage had been introduced, *and* moderators had been trained to use it after controlling for all other factors in the model.

For *red* messages, the model was again significant (χ^2_{22}=372; N=6264, $P<.001$), and explained between 5.8% (Cox and Snell R^2) and 8.6 % (Nagelkerke R^2) of the variance in moderator response and correctly classified 76.2% of messages. The strongest predictor of moderator response was the number of active moderators at the time the message was posted, recording an OR of 1.5. This indicated that messages that were posted when there were more moderators online were slightly more likely to receive a response, after controlling for all other factors in the model. The second strongest predictor of moderator response was whether training for the triage system had been conducted (OR 1.3).

Finally, for *crisis* messages, the model was also significant (χ^2_{22}=76.3; N=739, $P<.001$). It explained between 9.8% (Cox and Snell R^2) and 15.8% (Nagelkerke R^2) of the variance and correctly classified 81.2% of messages. The strongest predictor was the number of active moderators at the time the message was posted, recording an OR of 1.7, followed by the number of community posts at the time the message was posted. These were the only 2 variables to contribute significantly to the regression model. This indicates that crisis messages posted when there were more active moderators were almost twice as likely to receive a response after controlling for all other factors in the model. Neither the presence of the triage system nor training made statistically significant contributions to the model ($P>.1$).

Table 3. Proportion of messages that receive a reply from a moderator.

Priority	Prelaunch, n (%)	Postlaunch, n (%)	Post-training, n (%)
Green	17,004 (33.6)	1687 (21.6)	9699 (44.1)
Amber	5763 (65.2)	976 (59.7)	5322 (60.2)
Red	2568 (75.8)	461 (65.5)	1739 (80.1)
Crisis	308 (80.6)	30 (81.1)	261 (81.6)

Table 4. Time taken for moderators to respond to messages.

Priority	Prelaunch, median (IQR[a])	Postlaunch, median (IQR)	Post-training, median (IQR)
Green	0:37:38 (10:24:35)	0:21:49 (6:54:12)	0:33:16 (10:35:09)
Amber	2:08:34 (11:33:11)	0:33:22 (4:05:39)	0:30:08 (4:48:54)
Red	2:09:45 (9:33:11)	0:42:17 (4:57:47)	0:26:25 (1:52:00)
Crisis	2:23:03 (7:23:54)	0:47:11 (1:40:34)	0:28:45 (1:15:50)

[a] IQR: interquartile range.

Time Taken for Messages to Receive a Reply From a Moderator

Between prelaunch and postlaunch periods, the median time taken for moderators to respond to crisis messages was reduced from 2 hours 23 min to 47 min (see Table 4). Reductions were also observed for red (2 hours 10 min to 42 min), amber (2 hours 9 min to 33 min), and green messages (from 38 min to 22 min). Response latency also decreased between the postlaunch and post-training periods for crisis (from 47 min to 29 min), red (from 42 min to 26 min), and amber (from 33 min to 30 min) but increased for green messages (22 min to 33 min). Cumulatively, the presence of the triage system and the training of moderators resulted in reductions of 80%, 80%, 77%, and 12% for crisis, red, amber, and green messages, respectively.

Furthermore, the triage system reduced the variability of response latency for nongreen messages. The interquartile range (IQR) of latency for crisis, red, and amber messages was reduced by 77%, 48%, and 65%, respectively, between prelaunch and postlaunch periods. Between prelaunch and post-training periods, the same measures were reduced by 83%, 80%, and 58%, respectively. The IQR of latency for green messages decreased by 34% between prelaunch and postlaunch but remained steady (a 2% increase) between prelaunch and post-training.

A multiple regression analysis was used to further evaluate whether these differences in latency were statistically significant and whether they remained after accounting for moderator and community activity at and around the time of the message post (note: adjusted P=.004). Separate regression models were conducted for each priority level and included the same variables as the models for response ratio (described above).

For *green* messages, the total variance explained by the model as a whole was 47.9% ($F_{21,28368}$=1243.7; P<.001). In the final model, the number of moderators online at the time the message was posted was the strongest predictor of moderator response latency (beta=−.33; P<.001), indicating that the more moderators online, the shorter the latency in response time. Whether or not moderators had been trained to use the triage system also contributed significantly to the model indicating that response time reduced significantly from prelaunch to postlaunch periods (beta=−.09; P<.001).

For *amber* messages, the total variance explained by the model as a whole was 18.8% ($F_{21,12039}$=132.4; P<.001). Moderator activity during the hours before, at, and after a community member posted to the website, as well as triage period, were entered into the model. The number of active moderators at the time of the message made the strongest significant contribution to explaining moderator response latency (beta=−.35; P<.001). The presence of the triage system also made significant contribution to the model (beta=−.08; P<.001) indicating that response latency decreased from the prelaunch to postlaunch periods.

For *red* messages, the total variance explained by the model as a whole was 24.9% ($F_{21,4746}$=76.2; P<.001). Moderator activity during the hours before, at, and after a community member posted to the website, as well as trial period, were entered into the model. In the final model, the number of active moderators at the time of the message made the strongest significant contribution (beta=−.40; P<.001), indicating that the more moderators online when a community member posted, the shorter the latency in response time. The presence of the triage system also made a statistically significant contribution (beta=−.14; P=.001) with response latency significantly decreasing from prelaunch to postlaunch periods.

For *crisis* messages, the total variance explained by the final model as a whole was 28.2% ($F_{21,577}$=12.2; P<.001). The model consisted of moderator and community activity in the hours before, at, and after each message was posted. The strongest predictors of response latency were the number of active moderators at the time of the message (beta=−.34; P=.001), the number of moderator posts (beta=−.26; P=.001), and the number of active moderators in the hour preceding the message (beta=−.21; P=.003). This indicates that greater numbers of active moderators correspond to faster replies. Finally, the trial period also made a statistically significant contribution to the model (beta=−.17; P=.001).

To look more closely at the effect of triage period, planned comparisons with statistical correction (adjusted alpha=.004) comparing prelaunch and postlaunch periods (postlaunch and post-training) were conducted. These showed that messages were responded to more quickly during the post-training period (ie, combining the triage system with appropriate training) if they were labelled amber (P<.001) or red (P<.001). The large apparent difference in response latencies for crisis messages was significant only at the trend level (P=.007), likely because there were only 30 crisis messages in the postlaunch period.

Similar comparisons between postlaunch and post-training periods showed that messages were responded to more slowly in the post-training period if they were labelled green (P<.001) but more quickly if they were labelled red (P<.001), but the differences for amber and crisis messages were not significant (P>.05).

Finally, comparisons between prelaunch and post-training periods (ie, combining the triage system with appropriate training) showed that messages posted during the later period received replies significantly faster for all severity labels (*P*<.001).

Discussion

Summary of Findings

This study evaluated a triage system in terms of the accuracy with which it automatically prioritized content in online peer support and the extent to which it improved the responsiveness of human moderators to the prioritized content. The triage algorithm achieved high accuracy (84% f-measure) in identifying content requiring moderator response. Additionally, the combination of the triage system and appropriate training resulted in modest improvements to response ratio for all priority levels other than amber and large reductions to response latency for all priority levels other than green. Overall, the observed reductions in response latency and variability of response latency for flagged messages indicate that the triage system supported the online moderator as intended.

Accuracy of the Triage Algorithm

Over the holdout test set, the triage algorithm was more accurate in separating out flagged (ie, nongreen) messages than it was in separating urgent (ie, red and crisis) messages. Arguably, the first of these boundaries is more important because this determines which messages enter the sidebar (see Multimedia Appendix 1). A low recall here could cause the moderators to miss posts that they should pay attention and respond to, whereas a low precision would increase their workload by filling the sidebar with low-priority messages. In contrast, mistakes made in separating urgent and nonurgent messages only effect the ranking of messages within this sidebar.

In addition to the above results, it was encouraging to observe a high level of agreement between the moderators and the triage algorithm during the prelaunch period (ie, the first column of Table 3), where there was a clear progression in which the likelihood of moderators responding to messages was proportional to the priority assigned by the algorithm. This agreement was in no way due to the triage algorithm influencing moderator behavior as, during severity, labels were assigned retroactively for the prelaunch period (ie, after the moderators' responses had been made). Conversely, it was also in no way due to the moderators influencing the algorithm, as none of the features or training data used by the algorithm were based on moderator behavior. Thus, we are able to show that the moderators and the triage algorithm arrived at similar decisions independently.

Impact on Moderator Response Ratio

As mentioned previously, the triage system resulted in only modest improvements to moderator response ratio. This is understandable, given that the moderators already prioritized responding to urgent messages before the introduction of the triage system. Evidence of this response hierarchy is seen in the clear progression within the prelaunch period (ie, the first column of Table 3), where crisis messages were more likely to receive replies than red messages which were responded to more often than amber messages which were in turn more likely to receive a response than green messages. The potential to improve response patterns after launch and/or training was limited because the moderators were already behaving as desired.

In fact, the introduction of the triage system appears to have initially had a detrimental effect, with response ratios dropping between prelaunch and postlaunch periods for all severity labels other than crisis. This may be because of the unfamiliarity of the system and the lack of training given during this initial postlaunch period. Fortunately, response ratios recovered in the post-training period, such that the end result (ie, between prelaunch and post-training periods) was an increase across all severity labels other than amber. The reduction in response ratio for amber messages is likely due to the way the triage interface allows these messages to be clearly marked as resolved when the community has rallied around them. In the post-training period, moderators were specifically instructed to only respond to amber messages if they had been overlooked by the community.

The strongest predictor for response ratio was whether or not moderators had been trained to use the triage system (for green and amber posts) or the number of moderators active at the time of the message (for red and crisis). It is important to note that the introduction of the triage system (with training) significantly increased the likelihood that green, amber, and red messages received a moderator response after accounting for moderator and community activity. The system was not a significant predictor for the response ratios of crisis messages.

Impact on Moderator Response Latency

Before the introduction of the triage system, moderators took a median time of roughly 2 hours to respond to nongreen messages, whereas green messages tended to be responded to either quickly or not at all. The informal launch of the triage system led to large reductions in response latencies for amber, red, and crisis messages and modest reductions for green messages. The formal launch and associated training led to further reductions for amber, red, and crisis messages, and a substantial increase for green messages. Cumulatively (ie, between prelaunch and post-training periods), response ratios dropped by approximately 80% for crisis, red, and amber messages and 12% for green messages. Additionally, the variability (ie, IQR) of response rates dropped by approximately 60% for amber messages and approximately 80% for red and crisis messages but rose by 2% for green messages.

Across all priority levels, the strongest predictor of response latency was the number of moderators online at or around the time of the post. The formal launch of the triage system coincided with an influx of new volunteers, and their introduction had a large impact on response latency. However, it is important to note that the largest decreases in response latencies occurred between the prelaunch and postlaunch periods (ie, before this influx of new moderators). Additionally, the triage system and/or moderators' training with it were shown to be a significant predictor for the reduced response latencies across all priority levels. Overall, the large reductions in response latency and variability of response latency for priority messages indicate that the triage system supports moderator behavior as intended.

Limitations

This evaluation of moderator behavior has made a key assumption that the decisions made by the triage algorithm are correct and that it is desirable that moderators follow its recommendations. As mentioned above, the evaluation over the CLPsych 2016 dataset and the agreement observed between the moderators and the algorithm during the prelaunch period are encouraging. Nevertheless, it is important that future research evaluates the impact of the mistakes that any automated system will inevitably make, particularly given the inherent subjectivity of the prioritization task.

False negatives—messages that the system does not prioritize highly enough—have the potential to be particularly problematic. As moderators come to rely on the triage system, they may neglect to look elsewhere for high-priority content and consequently the system may counter-productively increase the chance that it is overlooked. We employed 2 strategies to decrease the likelihood of false negatives. The first was to reweight the triage algorithm so that it prioritizes recall ahead of precision, that is, it cares more about ensuring that any potentially urgent messages are included in the sidebar than ensuring that the sidebar includes only urgent messages. The second strategy was to deploy a tool that allowed ordinary community members to manually identify urgent content and add it to the triage system. The interface that enabled this crowdsourcing is described at the end of Multimedia Appendix 1. We would encourage other practitioners to provide similar safety nets if they adopt or develop an automated triage system.

Another limitation is that the analysis focused exclusively on the behavior of moderators and has not considered replies and support offered by community members, affiliates, or moderators in training. Moderators may systematically avoid responding to messages that receive a strong response from the community or encourage the community to be more self-reliant by withholding intervention when it is safe to do so. It is also possible that moderators will direct trainees to respond to urgent content rather than resolving it themselves. Our analysis has—necessarily, for the purposes of triage evaluation—assumed that a message is not resolved until it receives a reply from a moderator and that prompt replies are universally desirable. It would be more accurate to say that moderators should be kept aware of concerning content and be ready to intervene if necessary, rather than that they should always intervene as quickly as possible.

A related limitation is that the analysis has focused on moderator behavior without evaluating the impact this has on the community and its members. Although the underlying aim of increasing moderator responsiveness to urgent content has been to improve the safety and therapeutic value of the online community, we have not measured such outcomes directly. For future studies, it will be extremely valuable to survey forum users before and after the introduction of a triage system such as this to assess whether the changes to moderator behavior were noticeable and whether this leads to better outcomes or an improved sense of support.

A final, more technical limitation is that some of the variables considered by our statistical models are not entirely independent. Some of the variables introduced when modelling both response ratio and latency relate to the activity levels of moderators at and around the time of a message being posted. Intuitively, greater numbers of active moderators are likely to lead to more prompt responses. Unfortunately, moderator activity levels are not entirely independent from response ratio or latency, as there has to be at least 1 active moderator during one of these windows for a message to receive a reply. Consequently, these models may overestimate the impact of moderator activity. The same variables are also not entirely independent from the presence of the triage system, as it is possible that it has contributed to the activity levels of moderators by making the process of moderation more engaging. The combination of these 2 issues means that it is possible that these models have underestimated the individual impact of the triage system.

Comparison With Prior Work

To our knowledge, this study is the first evaluation of automated triage in online peer support that has focused on behavior change, that is, the system's ability to influence human moderators and direct their attention to where it is most needed.

There is, however, a great deal of prior work related to the machine learning and computational linguistics aspects of this study. This includes the CLPsych 2016 shared task [20] in which 15 teams of researchers competed to develop the best algorithm for prioritizing forum messages from ReachOut.com. Our classification system was directly informed by the submissions of the top performing teams [21-23] and was trained and evaluated on the same dataset. Encouragingly, researchers have continued to work with these data and hone their algorithms, with Cohan et al [24] significantly outperforming all previous participants. Additionally, a second edition of the shared task was run in 2017 and attracted 15 teams of researchers [25]. Thus, advances to the algorithm have already been made and are available for integration in future deployments of the triage system.

Also closely related are reports from Huh et al [26] and Delort et al [27] that describe machine learning systems to determine whether or not a forum message requires moderation. Both differ from our research by framing the problem as binary (a message either requires moderation or does not), whereas we have framed it as a multiclass prioritization problem. Both focus exclusively on algorithm accuracy, and do not investigate moderator behavior change.

In addition, there is a great deal of more broadly related work on the use of machine learning to detect undesirable content and behavior online. For example, researchers have developed algorithms to detect hate speech [28,29], cyberbullying [30-32], and the grooming activities of pedophiles [33]. There has also been much progress recently in applying natural language processing to social media to gain insights into the authors' state of mind, or to identify and diagnose individuals who could benefit from some form of psychological intervention or assistance (see [34] for a recent review). For example, researchers have developed algorithms to detect suicide ideation [35-38], depression [39-41], post-traumatic stress disorder [41,42], and the *dark triad* of antisocial personality traits [43,44]. It is very likely that such algorithms could be applied successfully to triaging content in online peer support.

Conclusions

This study has described a triage system that automatically prioritizes content in online peer support to augment human moderators and help them focus their efforts on the individuals and messages that have the greatest need. Through evaluation on a dataset of manually prioritized forum messages, we have shown that the triage algorithm is largely accurate, particularly for the critical boundary separating content that moderators need to pay attention to from content they can safely leave for peers to address on their own. Through a long-term field study, we have shown that the triage system greatly reduces the time taken for moderators to respond to prioritized content, and that moderators are largely in agreement with the triage system about which messages they should prioritize responding to.

Our underlying aim of this study in improving the responsiveness and scalability of human moderation in online peer support is to increase the safety and therapeutic value of these communities. However, further evaluations are needed to establish whether this is the case. Additionally, in the near future, we plan to investigate the impact of mistakes made by the triage algorithm and how best to encourage moderators and peers to provide an additional safety net (and additional training data) through manual prioritization. In addition, there are many opportunities to incorporate related work in social media mining to improve the accuracy of the triage algorithm that we hope to explore.

Acknowledgments

We would like to thank ReachOut.com's moderators for their continued efforts to provide a safe online environment for young people, and for being receptive to the triage system. We would also like to thank Sophie Potter and Ben Davies at ReachOut.com for their input into the design of the triage system, and the many researchers who participated in the CLPsych 2015 shared task. This work was partially supported by the Australian Research Council LP13010045. RAC is supported by the Australian Research Council Future Fellowship FT140100824. KLM was supported by UC Davis Autism Research Training Program Award Number T32MH073124 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A description of the triage interface, including screenshot.

[DOCX File, 435KB - jmir_v21i4e11410_app1.docx]

Multimedia Appendix 2

A description of the triage algorithm, and the features it relies on.

[DOCX File, 15KB - jmir_v21i4e11410_app2.docx]

Multimedia Appendix 3

Additional details about how moderators were identified, how replies were tracked for messages, and how the activity histograms were constructed.

[DOCX File, 75KB - jmir_v21i4e11410_app3.docx]

References

1. Miyamoto Y, Sono T. Lessons from peer support among individuals with mental health difficulties: a review of the literature. Clin Pract Epidemiol Ment Health 2012;8:22-29 [FREE Full text] [doi: 10.2174/1745017901208010022] [Medline: 22563347]
2. Barak A, Dolev-Cohen M. Does activity level in online support groups for distressed adolescents determine emotional relief. Couns Psychother Res 2006 Sep;6(3):186-190. [doi: 10.1080/14733140600848203]
3. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Comput Human Behav 2008 Sep;24(5):1867-1883. [doi: 10.1016/j.chb.2008.02.004]
4. Tanis M. Health-related on-line forums: what's the big attraction? J Health Commun 2008;13(7):698-714. [doi: 10.1080/10810730802415316] [Medline: 18958781]

5. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
6. Feinstein BA, Bhatia V, Hershenberg R, Davila J. Another venue for problematic interpersonal behavior: the effects of depressive and anxious symptoms on social networking experiences. J Soc Clin Psychol 2012 Apr;31(4):356-382. [doi: 10.1521/jscp.2012.31.4.356]
7. Suler J. The psychology of text relationships. In: Kraus R, Stricker G, Speyer C, editors. Online Counseling: A Handbook for Mental Health Professionals. Cambridge, MA: Academic Press; 2010:21-53.
8. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
9. Whitlock J, Lader W, Conterio K. The internet and self-injury: what psychotherapists should know. J Clin Psychol 2007 Nov;63(11):1135-1143. [doi: 10.1002/jclp.20420] [Medline: 17932984]
10. Lawlor A, Kirakowski J. Online support groups for mental health: a space for challenging self-stigma or a means of social avoidance? Comput Human Behav 2014 Mar;32:152-161. [doi: 10.1016/j.chb.2013.11.015]
11. Chang T. Online counseling: prioritizing psychoeducation, self-help, and mutual help for counseling psychology research and practice. Couns Psychol 2016 Jun 30;33(6):881-890. [doi: 10.1177/0011000005279962]
12. Webb M, Burns J, Collin P. Providing online support for young people with mental health difficulties: challenges and opportunities explored. Early Interv Psychiatry 2008 May;2(2):108-113. [doi: 10.1111/j.1751-7893.2008.00066.x] [Medline: 21352141]
13. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: designing and evaluating technology for mental health. ACM Trans Comput-Hum Interact 2014 Feb 01;21(1):1-26. [doi: 10.1145/2513179]
14. Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]
15. Wise K, Hamman B, Thorson K. Moderation, response rate, and message interactivity: features of online communities and their effects on intent to participate. J Comp Mediated Comm 2006 Oct;12(1):24-41. [doi: 10.1111/j.1083-6101.2006.00313.x]
16. Chen J, Xu H, Whinston AB. Moderated online communities and quality of user-generated content. J Manag Inf Syst 2011 Nov 14;28(2):237-268. [doi: 10.2753/mis0742-1222280209]
17. Burns J, Morey C, Lagelée A, Mackenzie A, Nicholas J. Reach Out! Innovation in service delivery. Med J Aust 2007 Oct 1;187(7 Suppl):S31-S34. [Medline: 17908022]
18. Reachout Australia. Reachout. 2015. Reachout Australia Annual Report 2015-2016 URL: http://about.au.reachout.com/wp-content/uploads/2016/12/ReachOut-2015-16-Annual-Report_pages.pdf [accessed 2019-02-05] [WebCite Cache ID 75xA4DjnF]
19. Metcalf A, Blake V. ReachOut. 2013. 2013 ReachOut Annual User Survey Results URL: http://about.au.reachout.com/wp-content/uploads/2015/01/ReachOut.com-Annual-User-Survey-2013.pdf [accessed 2019-02-05] [WebCite Cache ID 75xAGrJV2]
20. Milne DN, Pink G, Hachey B, Calvo RA. Shared Task: Triaging content in online peer-support forums. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: CLPsych 2016 Computational Linguistics and Clinical Psychology; June 16, 2016; San Diego, CA p. 118-127.
21. Mac Kim S, Wang Y, Wan S, Paris C. Data61-CSIRO systems at the CLPsych 2016 Shared Task. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: Computational Linguistics and Clinical Psychology; June 2016; San Diego, CA p. 128-132. [doi: 10.18653/v1/W16-0313]
22. Malmasi S, Zampieri M, Dras M. Predicting Post Severity in Mental Health Forums. In: Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology. 2016 Presented at: Computational Linguistics and Clinical Psychology; June, 2016; San Diego, CA p. 133-137. [doi: 10.18653/v1/W16-0314]
23. Brew C. Classifying ReachOut posts with a radial basis function SVM. In: Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology. 2016 Presented at: Computational Linguistics and Clinical Psychology; June 2016; San Diego, CA p. 138-142.
24. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. J Assoc Inf Sci Technol 2017 Sep 25;68(11):2675-2689. [doi: 10.1002/asi.23865]
25. Milne DN. CLPsych. 2017. Triaging content in online peer-support: an overview of the 2017 CLPsych shared task URL: http://clpsych.org/shared-task-2017/ [accessed 2018-01-05] [WebCite Cache ID 6wEjBpCcp]
26. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013 Dec;46(6):998-1005 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.011] [Medline: 24025513]
27. Delort JY, Arunasalam B, Paris C. Automatic moderation of online discussion sites. Int J Electron Commer 2011 Apr 26;15(3):9-30. [doi: 10.2753/JEC1086-4415150302]
28. Nobata C, Tetreault J, Thomas A, Mehdad Y, Chang Y. Abusive Language Detection in Online User Content. In: Proceedings of the 25th International Conference on World Wide Web. 2016 Presented at: WWW '16; 2016; Montreal, Canada p. 145-153. [doi: 10.1145/2872427.2883062]

29. Djuric N, Zhou J, Morris R, Grbovic M, Radosavljevic V, Bhamidipati N. Hate Speech Detection with Comment Embeddings. In: WWW '15 Companion. 2015 Presented at: 24th International Conference on World Wide Web; May 18-22, 2015; Florence, Italy p. 29-30. [doi: 10.1145/2740908.2742760]
30. Al-garadi MA, Varathan KD, Ravana SD. Cybercrime detection in online communications: the experimental case of cyberbullying detection in the Twitter network. Comput Hum Behav 2016 Oct;63:433-443. [doi: 10.1016/j.chb.2016.05.051]
31. Dadvar M, de Jong F. Cyberbullying detection: a step toward a safer internet yard. In: WWW '12 Companion. 2012 Presented at: 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 121-126. [doi: 10.1145/2187980.2187995]
32. Zhao R, Zhou A, Mao K. Automatic detection of cyberbullying on social networks based on bullying features. In: ICDCN '16. 2016 Presented at: 17th International Conference on Distributed Computing and Networking; January 04-07, 2016; Singapore, Singapore. [doi: 10.1145/2833312.2849567]
33. Cano AE, Fernandez M, Alani H. Detecting child grooming behaviour patterns on social media. 2014 Presented at: International Conference on Social Informatics; 2014; Barcelona, Spain p. 412-427. [doi: 10.1007/978-3-319-13734-6_30]
34. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30;23(5):1-37. [doi: 10.1017/S1351324916000383]
35. Wood A, Shiffman J, Leary R, Coppersmith G. Language signals preceding suicide attempts. 2016 Presented at: CHI Workshop on Computing and Mental Health; 2016; San Jose, CA.
36. Homan CM, Johar R, Liu T, Lytle M, Silenzio V, Alm C. Toward macro-insights for suicide prevention: analyzing fine-grained distress at scale. 2014 Presented at: Computational Linguistics and Clinical Psychology; 2014; Baltimore, MD p. 107-117.
37. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
38. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. 2016 May Presented at: CHI Conference on Human Factors in Computing Systems; 2016; San Jose, CA p. 2098-2110 URL: http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]
39. De Choudhury M, Gamon M. Predicting depression via social media. 2013 Presented at: Weblogs and Social Media; 2013; Ann Arbor, MI p. 128-137.
40. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: CHI '13. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 02, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
41. Coppersmith G, Dredze MH, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: Computational Linguistics and Clinical Psychology; June 5, 2015; Denver, CO p. 31-39.
42. Coppersmith G, Dredze MH, Harman C. Measuring post traumatic stress disorder in Twitter. 2014 Presented at: Weblogs and Social Media; 2014; Ann Arbor, MI p. 579-582.
43. Sumner C, Byers A, Boochever R, Park G. Predicting dark triad personality traits from twitter usage: a linguistic analysis of tweets. In: 11th International Conference on Machine Learning and Applications. 2012 Presented at: ICMLA '12; 2012; Boca Raton, FL p. 386-393. [doi: 10.1109/ICMLA.2012.218]
44. Garcia D, Sikström S. The dark side of Facebook: semantic representations of status updates predict the Dark Triad of personality. Pers Individ Dif 2014 Sep;67:92-96. [doi: 10.1016/j.paid.2013.10.001]

Abbreviations

CLPsych: Computational Linguistics and Clinical Psychology
IQR: interquartile range
OR: odds ratio

Edited by G Eysenbach; submitted 12.07.18; peer-reviewed by B Polepalli Ramesh, K Machin; comments to author 14.10.18; revised version received 22.11.18; accepted 09.12.18; published 26.04.19.

Please cite as:
Milne DN, McCabe KL, Calvo RA
Improving Moderator Responsiveness in Online Peer Support Through Automated Triage
J Med Internet Res 2019;21(4):e11410
URL: https://www.jmir.org/2019/4/e11410/
doi:10.2196/11410
PMID:31025945

Original Paper

A Stroke Risk Detection: Improving Hybrid Feature Selection Method

Yonglai Zhang[1*], PhD; Yaojian Zhou[1*], PhD; Dongsong Zhang[2*], PhD; Wenai Song[1*], PhD

[1]Medical Big Data Institute, Software School, North University of China, Taiyuan, China

[2]Department of Business Information Systems and Operations Research, Belk School of Business, University of North Carolina, Charlotte, NC, United States

[*]all authors contributed equally

Corresponding Author:
Yaojian Zhou, PhD
Medical Big Data Institute
Software School
North University of China
3 Xueyuan St
Taiyuan, 030051
China
Phone: 86 3923450
Email: zhouyj@nuc.edu.cn

Abstract

Background: Stroke is one of the most common diseases that cause mortality. Detecting the risk of stroke for individuals is critical yet challenging because of a large number of risk factors for stroke.

Objective: This study aimed to address the limitation of ineffective feature selection in existing research on stroke risk detection. We have proposed a new feature selection method called weighting- and ranking-based hybrid feature selection (WRHFS) to select important risk factors for detecting ischemic stroke.

Methods: WRHFS integrates the strengths of various filter algorithms by following the principle of a wrapper approach. We employed a variety of filter-based feature selection models as the candidate set, including standard deviation, Pearson correlation coefficient, Fisher score, information gain, Relief algorithm, and chi-square test and used sensitivity, specificity, accuracy, and Youden index as performance metrics to evaluate the proposed method.

Results: This study chose 792 samples from the electronic records of 13,421 patients in a community hospital. Each sample included 28 features (24 blood test features and 4 demographic features). The results of evaluation showed that the proposed method selected 9 important features out of the original 28 features and significantly outperformed baseline methods. Their cumulative contribution was 0.51. The WRHFS method achieved a sensitivity of 82.7% (329/398), specificity of 80.4% (317/394), classification accuracy of 81.5% (645/792), and Youden index of 0.63 using only the top 9 features. We have also presented a chart for visualizing the risk of having ischemic strokes.

Conclusions: This study has proposed, developed, and evaluated a new feature selection method for identifying the most important features for building effective and parsimonious models for stroke risk detection. The findings of this research provide several novel research contributions and practical implications.

(J Med Internet Res 2019;21(4):e12437) doi:10.2196/12437

KEYWORDS

machine learning; stroke; risk; feature selection; WRHFS

Introduction

Background and Research Objective

Stroke is the second most popular cardiovascular disease (CVD). The World Health Organization estimated that 17.7 million people died from CVDs in 2017, of which 6.7 million had stroke, representing 31% of all deaths caused by diseases in the world [1]. The epidemiological characteristics of stroke in developing countries have gradually become closer to those of developed countries [2]. The prevalence and mortality of stroke are still on the rise. As of 2016, there were 13 million people with stroke in China [3]. Stroke prevention was the theme set by the World Stroke Organization for the 2017 World Stroke Day. Therefore, timely detection and prevention of stroke become essential.

People may go to a hospital for a full physical examination to assess stroke risk. Specific examination items include blood biochemical tests, blood pressure, electrocardiogram, vascular ultrasound, vascular computerized tomography angiography, magnetic resonance angiography, electroencephalography, magneto encephalography, single photon emission computerized tomography, positron emission computerized tomography, magnetic resonance imaging, and digital subtraction angiography. Plaque image analysis based on image segmentation technology has also been explored for risk detection of strokes [4,5].

As traditional medical risk assessment is expensive and not scalable, automated detection of stroke risk has been increasingly studied in recent years (eg, [6-11]), which falls into 2 broad categories: stroke risk assessment modeling and brain image analysis. Many countries have employed automated detection models for stroke, such as systematic coronary risk evaluation [12], QRISK (QRFSEARCH cardiovascular risk algorithm) [13], and Reynolds risk score [14]. The Framingham risk assessment model is a typical risk detection model of stroke. Pencina et al used an extended Framingham model to develop a 30-year risk detection model with data collected from 4506 patients aged 20 to 59 years [15]. The model detected the 30-year risk using 8 risk factors, including gender, antihypertensives, blood pressure, total cholesterol, high-density lipoprotein (HDL), smoking, impaired glucose tolerance, and left ventricular hypertrophy. Flueckiger et al further extended the Framingham model to establish a score detection model of stroke in a multiethnic study of atherosclerosis in conjunction with nontraditional risk markers [16]. The detection model included demographics, medical history, anthropometrics, and conventional risk factors. However, the Framingham model overestimates the risk of stroke in China because of obvious differences in the disease spectrum and risk factors [17,18]. A joint Chinese-American research group constructed a risk detection model of ischemic stroke and hemorrhagic stroke with 6 risk factors, including systolic blood pressure, sex, age, total cholesterol (TC), diabetes, and smoking [19].

Stroke consists of ischemic stroke and hemorrhagic stroke. Ischemic stroke accounts for 60% to 80% of stroke occurrence in China, which is the main context of this study. The detection of risk for ischemic stroke is aimed to reduce or prevent the incidence of clinical events and premature death associated with ischemic stroke by early prevention. A key limitation of existing research on stroke risk assessment lies in the lack of systematic guidance for feature selection while building stroke risk detection models, which is essential to the performance of such models. Previous studies chose predictive features largely in an ad hoc manner and did not incorporate the latest results of medical research. So, the core research question of this study is how to select important risk factors that should be included in a risk detection model for ischemic stroke as predictive features?

To address this research question, we proposed, developed, and evaluated a new hybrid feature selection method, namely weighting- and ranking-based hybrid feature selection (WRHFS). WRHFS integrates the strengths of various filter algorithms and deploys continuous weighting and ranking of individual features by following the principle of a wrapper approach. It then selects the top N ranked features as the most important features. This study makes a significant research contribution by proposing a new methodological approach to feature selection, which can lead to improved performance of risk detection models.

Related Work

The key to accurate stroke risk detection is to select the most important and influential features of stroke patients, which may vary among patients at different regions.

Past research has shown that stroke is significantly associated with age [20], gender [21], blood pressure [20,21], low-density lipoprotein [22], triglyceride [23], drinking [24], smoking [25], creatine kinase (CK) [25], height [26], TC [27], HDL [24,27], body mass index (BMI) [22,25,28], serum total cholesterol [22,29], smoking [22,24,30], and diabetes [22,31]. Recently, some new risk factors have been discovered by medical research. For example, alkaline phosphatase [32] and hypercholesterolemia [33] are found to increase the probability of the mortality of stroke patients. Studies have also shown that there is a clear epidemiological relationship between stroke risk and hyperlipidemia [34]. However, no single study has used all features that are theoretically related to stroke because of their availability in data.

Traditionally, detectors of stroke risk were identified based on the findings of medical research and practice. However, collecting data for risk factors (also referred to as features in this paper) based on the results of medical research is extremely difficult. In the past decade, there has been increasing research on building automated stroke risk detection models by leveraging machine-learning techniques and patient data. One of the essential steps in building such models is to select effective features (ie, influential factors) that are associated with stroke, which is often referred to as the feature selection process. We categorized feature selection methods used in automated stroke risk detection models into semisupervised, unsupervised, and supervised methods [35,36], as summarized in Table 1. Semisupervised feature selection methods are suitable for datasets with a small number of labeled samples and a large number of unlabeled samples [37]. The key challenge lies in how to use the labeled samples to efficiently process the unlabeled samples. At present, unsupervised feature selection methods mainly focus on clustering-based models, for example, Laplacian score [38], trace ratio [39], and sparsity regularization–based models [40]. For example, a coregularized unsupervised feature selection algorithm was proposed in a study by Zhu et al [41], which was intended to ensure that the selected features could preserve both data distribution and reconstruction.

Table 1. Classification of feature selection methods.

Methods	Rationale	Limitations	Sample studies
Supervised			
Filter	Mutual information based	Signal objective function	[42]
	Ranking based	Neglecting the correlation between the features and class labels	[43]
	Weighting based	Lacking the uniform standards of selecting features	[44]
Wrapper	Evaluating the accuracy of the classifier	Overfitting and high computational complexity	[45]
Hybrid	Guiding the wrapper using a filter	Only for certain specific fields	[46]
Semisupervised	Guiding by the labeled samples	Relying on small labeled samples	[37]
Unsupervised	Clustering-based models	Relying on certain data distribution	[40]

Supervised feature selection methods can be further divided into filter, wrapper, and hybrid methods. The filter feature selection method consists of mutual information and ranking- and weighting-based methods. Mutual information–based filter methods use mutual information to evaluate the relevance of features to class labels and the redundancy of candidate features. However, they suffer from the problem that the objective function only uses a single statistic measure of a dataset (eg, standard deviation, information gain [42,47], or Fisher score [48]), while ignoring the fusion of multiple measures. For example, a standard deviation–based filter model relies on the distance between feature value and mean value for feature selection. Information entropy is often used to measure the uncertainty of the value of a random variable. Information gain, referred to as the change in information entropy, of a feature in a dataset can be used to rank features. The greater the information gain is, the more a feature contributes to classification.

Feature ranking methods (eg, maximal relevance and minimal redundancy objective [43]) are independent of classification algorithms. They select a feature subset with metrics such as the Relief algorithm [49-51] and correlation estimate [43,52]. The Relief algorithm has been successfully applied to feature weighting because of its simplicity and effectiveness [41,42,47]. It is inspired by instance-based learning algorithms according to their ability to discriminate neighboring patterns. Linear in-time complexity, Relief has a great advantage in computational efficiency. It selects a sample x randomly and then finds the nearest neighbor sample NearHit(x) in the same class and the nearest neighbor sample NearMiss(x) in another class. However, its significant disadvantage lies in that feature ranking overemphasizes the relevance of a certain feature to a class label or the correlation with other individual features based on a single objective function, while neglecting the correlation between the combined features and a class label. In addition, when the independent relevance of a feature is emphasized, the redundancy of feature ranking will be increased, which contradicts to the objective of minimization of redundancy.

Feature weighting methods attempt to assign a weight value, usually in the range of 0 to 1, to each feature. Features with weights near 1 will be selected to form a feature set, whereas other features will be discarded [44]. Those methods are lacking the uniform standards for selecting features because of the fuzziness of *near 1*. Overall, filter models select features by weighting and ranking features based on their statistical relevance to class labels and a threshold to filter out irrelevant features to improve the classification accuracy [53].

Wrapper methods search for the optimal subset of features in a feature space and use a classifier to evaluate the effectiveness of a feature subset. For a particular classifier, wrapper methods may find good feature subsets [45]. However, they are prone to overfitting and high computational complexity.

Hybrid models use a filter model to guide a wrapper model to solve these problems of filter and wrapper methods [46,54-56]. In summary, for stroke risk detection, traditional feature selection methods have a variety of limitations, negatively affecting the quality of selected features and the performance of stroke risk detection models.

Methods

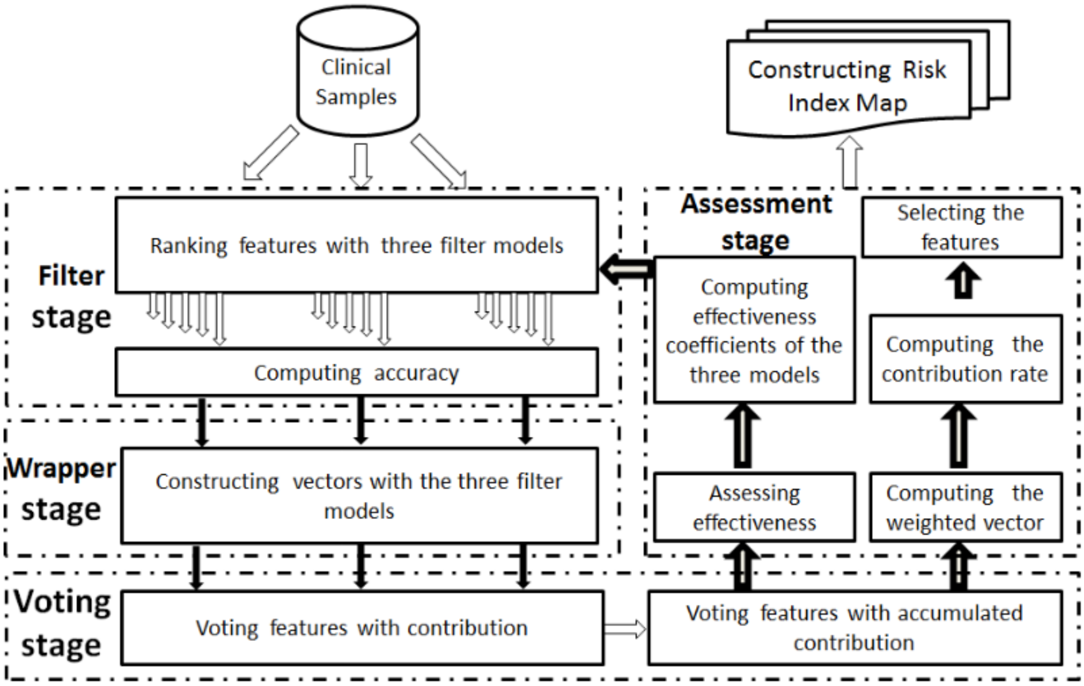
Design

In this study, we proposed a new hybrid feature selection model called WRHFS, which selects features by integrating various filter and wrapper methods. Being different from previous hybrid methods, WRHFS selects the best n filter models (in this study, n=3) from a candidate set to guide a wrapper model. Figure 1 shows the process of WRHFS, which consists of 4 parts. WRHFS selects the top 3 filter methods from a set of candidate filter models.

- Filter stage: ranking features with multiple filter models.

(1) Randomly choosing 3 different models from a set of candidate filter models.

Figure 1. Weighting- and ranking-based hybrid feature selection.



(2) Ranking features based on each filter model. WRHFS uses the ordered features of the filter models to train multiple classification models based on the backward searching strategy and measures classification accuracies and contribution vectors ω from the 3 filter models.

- Wrapper stage: constructing an aggregated contribution vector W using the 3 contributions of individual features from the 3 filter models.

(3) Creating W by aggregating the 3 contribution vectors ω_i (i=1, 2, and 3) generated by the 3 filter models. W is expressed as follows:

$W=\sum\omega_i$, i=1, 2, 3 (1)

- Voting stage: voting features based on a contribution matrix.

(4) Building a classification contribution matrix C based on the 3 contribution vectors. C is expressed as follows:

$C=[\omega1\omega2\omega3]$ (2)

(5) Building a cumulative classification contribution matrix D on the 3 contribution vectors ω_i. D_1, D_2, and D_3 are the cumulative contribution vectors based on the vectors ω_i, respectively. D is expressed as follows:

$D=[D_1D_2D_3]$ (3)

- Assessment stage: assessing the effectiveness of the 3 filter feature selection models and selecting the most important features.

(6) Building the effectiveness coefficient vector P of the 3 models and assessing the effectiveness of those models. P is a 3-dimensional vector, which contains the effectiveness coefficients of the 3 filter models. P is defined as follows:

$P=W \times D$ (4)

WRHFS assesses the first 3 filter models according to the effectiveness coefficient vector P, then replaces the worst filter model by choosing a filter model from the remaining filter models in the candidate set. Then, it repeats steps (1) to (6) until the candidate model set is empty. Finally, the top 3 filter models will be chosen based on the effectiveness coefficient vector P to develop the optimal feature selection model.

(7) Calculating the weight W_r of each selected individual feature. The formula of W_r is as follows:

$W_r= P \times C^T$(5)

(8) Ranking features based on the weights W_r.

(9) Selecting the top N key features based on their weights, in which the cumulative contribution of key features is more than 50% and generating the risk index map for diseases using the surface fitting technique based on key features.

Performance Measures

We evaluated the performance of WRHFS in terms of sensitivity, specificity, accuracy, and Youden index using a real-world dataset. We adopted the most common performance measures of classification models in medical diagnostics, including sensitivity, specificity, accuracy, and Youden index. There are 4 categories of potential outcomes: true positive (people with ischemic risk correctly identified), false positive (healthy people incorrectly identified as having risk), true negative (healthy people correctly identified as healthy), and false negative (people with ischemic stroke incorrectly identified as without risk). Sensitivity (also called the true positive rate or recall) measures the proportion of actual positives correctly detected as people with stroke risk, as shown in equation (6). Specificity (also called the true negative rate) measures the proportion of actual negatives that are correctly identified as healthy people, as shown in equation (7). Accuracy is defined as equation (8). Among the 3 measures, sensitivity is the most important medical criterion. Youden index, also called Youden J statistic, captures the performance of a dichotomous diagnostic test. Youden index is defined in equation (9).

Sensitivity = True positives / (True positives + False negatives) = True positives / Sick individuals (6)

Specificity = True negatives / (True negatives + False positives) = True negatives / Well individuals (7)

Accuracy = (True positives + True negatives) / All = True individuals / All individuals (8)

Youden index=Sensitivity + Specificity – 1 (9)

We employed 6 filter methods commonly used in the medical field, including those based on standard deviation [57], Pearson correlation coefficient [58], Fisher score [59], information gain [60], Relief [61], and chi-squared test [62]. We used a 10-fold cross-validation to train and test classification models. In the evaluation, we selected methods based on standard deviation, Pearson correlation coefficient, and Fisher score, initially. We adopted support vector machine (SVM), Bayes [63], classification based on associations [64], back-propagation neural networks [65], classification and regression tree [66], C4.5 (the decision tree learner) [67], and extreme learning machine [68] to build different detection models because they are the commonly used classification algorithms. Afterward, we kept the top 3 filter methods as the benchmark feature selection models and compared their performances against that of the proposed WRHFS method.

Results

Dataset

This study adopted a retrospective cohort. We collected a dataset that consisted of records of 80,672 patients from a community hospital. Among them, 13,421 patients suffered from ischemic stroke in the past 5 years. We extracted their records before their diagnoses of ischemic stroke. Given the purpose of modeling, we only chose and used features that did not have missing values in the entire dataset. We did not use missing value supplementation techniques because of concerns of possible biases or noises that may incur when applying those techniques.

At the end, there were 792 complete records in the dataset, each including 24 blood test features, as shown in Table 2. We also included 4 demographic features of the patients, including gender, age, height, and BMI. Descriptive statistics of age and gender are reported in Table 3. Among the 792 qualified patient records, 398 were diagnosed with ischemic stroke and labeled as class 1 instances, whereas the remaining 394 were not diagnosed with ischemic stroke and were labeled as class 2 instances.

Weighting of Features Using Weighting- and Ranking-Based Hybrid Feature Selection

WRHFS assessed the effectiveness of filter feature selection methods given in the dataset. Afterward, we discarded the filter method with lower effectiveness coefficients. The greater the coefficient, the higher the effectiveness. As shown in Table 4, information gain, Relief, and standard deviation led to the top 3 model performances. Table 5 shows the weights of the features based on standard deviation The weights of the features based on Relief and information gain are shown in Multimedia Appendices 1 and 2. Here, “accuracy” refers to the classification accuracy of the SVM classifier on the basis of the backward searching strategy, whereas “C” and “q” indicate the optimal penalty parameter and the kernel bandwidth in the SVM algorithm, respectively. Feature contribution is reflected by the difference between the accuracy of a model including a specific feature versus the accuracy of the model without it. We used the normalized result between 0 and 1 to eliminate the difference between positive and negative. “SD (0-1)” and “Contribution (0-1)” indicate the normalized results of “accuracy” and “contribution”, respectively. “Weight” reflects the overall performance of the features, which is the sum of standard deviation (0-1) and the contribution (0-1).

In Table 6, “weight sum” indicates the sum of the weights calculated by the 3 filter feature selection models. In Table 7, columns 2 to 4 compose the contribution matrix C, whereas columns 5 to 7 compose the cumulative contribution matrix D. In Table 8, the results of the weighted sum of the features using WRHFS are sorted in decreasing order, with a larger weight indicative of higher importance.

Table 9 presents the optimal performance of the trained risk detection models in terms of the 4 measures, including sensitivity (the positive detection rate), specificity (the negative detection rate), accuracy (the overall classification accuracy), and Youden index. From Table 9 it can be seen that the proposed WRHFS method achieved sensitivity of 82.7% (329/398) and classification accuracy of 81.5% (645/792) using only the top 9 features, and different classification models achieved the best performance when using different features. For example, information gain achieved the best classification accuracy of 72.5% (574/792) when using the top 10 features, and the accuracy began to decline when adding the eleventh feature. Similarly, standard deviation achieved the best classification accuracy of 73.2% (580/792) with the top 20 features presented in Table 5, and Relief achieved 72.9% (577/792) with the top 13 features. Therefore, we calculated sensitivity, specificity, accuracy, and Youden index of those methods by only using those optimal features that resulted in the best performed models. Among these feature selection methods, the proposed WRHFS method resulted in the highest performance measures with the fewest features. As shown in Table 8, Age, α-HBD, SCr, LDH, Height, TBIL, CK, Apo-B, and CK-MB are the top 9 most important features among the 28 features identified by WRHFS. Their cumulative contribution was 0.51. Table 10 presents the performances of models developed by different classifiers, as explained in Section “Performance Measures” using the same 9 features identified by WRHFS. Among all the models, SVM using WRHFS achieved the best performance in all 4 measures.

Table 2. 24 blood test items.

Full name	Abbreviation	Unit	Type of data
α Hydroxybutyric dehydrogenase	α- HBD	IU/L	Integer
Gamma glutamyl transpeptidase	GGP	IU/L	Integer
Lactate dehydrogenase	LDH	mmol/L	Real
Low-density lipoprotein	LDL	mmol/L	Real
High-density lipoprotein	HDL	mmol/L	Real
Blood urea nitrogen	BUN	mmol/L	Real
Uric acid	UA	umol/L	Integer
Total cholesterol	TC	mmol/L	Real
Total bilirubin	TBIL	umol/L	Real
Total protein	TP	g/L	Integer
Triglyceride	TG	mmol/L	Real
Albumin	Alb	g/L	Integer
Direct bilirubin	DBIL	umol/L	Real
Alkaline phosphatase	ALP	IU/L	Integer
Serum phosphorus	PI	mmol/L	Real
Serum creatinine	SCr	umol/L	Integer
Creatine kinase	CK	IU/L	Integer
Creatine kinase isoenzyme	CK-MB	IU/L	Integer
Glucose	Glu	mmol/L	Real
Alanine aminotransferase	ALT	IU/L	Integer
Aspartate aminotransferase	AST	IU/L	Integer
Apolipoprotein A1	Apo-A1	g/L	Real
Apolipoprotein B	Apo-B	g/L	Real
Serum calcium	Ca	mmol/L	Real

Table 3. Descriptive statistics of age and gender of patients in the dataset (N=792)

Age (years) and gender	Statistics, n (%)
≥ 45 and ≤60	
Male	105 (13.3)
Female	167 (21.1)
>60 and ≤75	
Male	151 (19.1)
Female	246 (31.1)
>75 and ≤90	
Male	76 (9.6)
Female	47 (5.9)

Table 4. Effectiveness coefficients of the filter feature selection methods.

Method	Effective coefficient
Information gain	63
Relief	61
Standard deviation	52
Pearson correlation coefficient	49
Fisher score	46
Chi-squared test	40

Table 5. Weighting of the 28 features based on standard deviation.

Feature[a]	Standard deviation	C	q	Accuracy (%)	Contribution	SD (0-1)	Contribution (0-1)	Weight
CK	0.21	16	0.5	56.1	—[b]	1.00	1.00	—
LDH	0.21	64	0.5	57.1	1.00	0.99	0.30	1.00
α-HBD	0.19	128	8.0	58.6	1.52	0.91	0.37	1.52
Height	0.17	8	8.0	57.3	−1.26	0.81	0.00	−1.26
ALP	0.15	2	4.0	58.8	1.52	0.72	0.37	1.52
UA	0.10	1	8.0	58.6	−0.25	0.48	0.13	−0.25
SCr	0.09	16	4.0	61.9	3.28	0.41	0.60	3.28
GGP	0.08	16	2.0	61.6	−0.25	0.40	0.13	−0.25
TP	0.08	2	8.0	61.6	0.00	0.37	0.17	0.00
AGE	0.08	64	1.0	67.9	6.31	0.36	1.00	6.31
ALT	0.07	128	1.0	67.6	−0.38	0.31	0.12	−0.38
AST	0.06	64	1.0	67.9	0.38	0.27	0.22	0.38
CK-MB	0.05	128	0.2	69.2	1.26	0.23	0.33	1.26
Alb	0.05	128	0.1	69.6	0.38	0.22	0.22	0.38
TBIL	0.04	256	0.5	72.1	2.53	0.16	0.50	2.53
BMI	0.03	64	0.3	72.7	0.63	0.12	0.25	0.63
Glu	0.01	128	0.3	72.7	0.00	0.04	0.17	0.00
DBIL	0.01	64	0.5	73.1	0.38	0.04	0.22	0.38
BUN	0.01	64	0.5	73.0	−0.13	0.03	0.15	−0.13
TC	0.01	64	0.5	73.2	0.25	0.02	0.20	0.25
LDL	0.01	128	1.0	73.0	−0.25	0.02	0.13	−0.25
TG	0.00	128	1.0	72.9	−0.13	0.02	0.15	−0.13
Gender	0.00	128	1.0	73.0	0.13	0.00	0.18	0.13
Ca	0.00	64	0.5	73.0	0.00	0.00	0.17	0.00
Apo-A1	0.00	128	1.0	73.2	0.25	0.00	0.20	0.25
HDL	0.00	128	1.0	73.1	−0.13	0.00	0.15	−0.13
Apo-B	0.00	128	1.0	73.1	0.00	0.00	0.17	0.00
PI	0.00	128	1.0	73.0	−0.13	0.00	0.15	−0.13

[a] The full forms of all abbreviations are shown in Table 2.

Table 6. Weighting of the 3 feature selection models.

Order	Feature[a]	Standard deviation	Relief	Information gain	Weight sum
1	α-HBD	0.9123	1.0000	0.0001	1.9124
2	GGP	0.4000	0.0657	0.0498	0.5156
3	Alb	0.2198	0.0592	0.0211	0.3001
4	LDL	0.0197	0.0026	0.0236	0.0459
5	TG	0.0156	0.0002	0.0001	0.0159
6	HDL	0.0032	0.0000	0.0010	0.0042
7	ALT	0.3120	0.0055	0.1141	0.4316
8	AST	0.2734	0.0366	0.0985	0.4085
9	SCr	0.4142	0.0637	0.0638	0.5417
10	CK	1.0000	0.5919	0.0549	1.6468
11	CK-MB	0.2303	0.0190	0.1657	0.4150
12	ALP	0.7239	0.0509	0.1051	0.8799
13	AGE	0.3574	0.0503	1.0000	1.4077
14	BUN	0.0296	0.0005	0.0845	0.1146
15	UA	0.4817	0.0024	0.0037	0.4878
16	LDH	0.9884	0.9582	0.0788	2.0254
17	Height	0.8145	0.4240	0.1235	1.3621
18	BMI	0.1171	0.0011	0.2146	0.3328
19	Gender	0.0049	0.0000	0.1349	0.1398
20	Ca	0.0040	0.0000	0.0000	0.0040
21	PI	0.0000	0.0001	0.0812	0.0813
22	Glu	0.0430	0.0009	0.4154	0.4593
23	Apo-A1	0.0036	0.0001	0.4525	0.4562
24	Apo-B	0.0013	0.0001	0.6987	0.7000
25	DBIL	0.0364	0.0003	0.2629	0.2996
26	TC	0.0248	0.0000	0.0382	0.0630
27	TBIL	0.1633	0.0323	0.5188	0.7143
28	TP	0.3667	0.0946	0.0417	0.5029

[a] The full forms of all abbreviations are shown in Table 2.

Table 7. Contribution of individual features.

Feature[a]	Contribution			Cumulative contribution		
	Standard deviation	Relief	Information gain	Standard deviation	Relief	Information gain
α-HBD	0.9123	1.0000	0.0001	1.6654	1.0000	6.2282
GGP	0.4000	0.0657	0.0498	2.8987	2.5000	5.1229
Alb	0.2198	0.0592	0.0211	4.9488	3.4428	5.8159
LDL	0.0197	0.0026	0.0236	6.5655	6.2714	5.5703
TG	0.0156	0.0002	0.0001	6.7155	7.8856	6.3685
HDL	0.0032	0.0000	0.0010	7.4155	9.4571	6.0966
ALT	0.3120	0.0055	0.1141	4.1821	6.0714	3.7369
AST	0.2734	0.0366	0.0985	4.3988	5.0000	4.0439
SCr	0.4142	0.0637	0.0638	2.7654	3.0857	4.8422
CK	1.0000	0.5919	0.0549	1.0000	1.6571	4.9562
CK-MB	0.2303	0.0190	0.1657	4.7321	5.8428	2.4474
ALP	0.7239	0.0509	0.1051	2.0320	3.6428	3.8158
AGE	0.3574	0.0503	1.0000	4.0654	4.6428	1.0000
BUN	0.0296	0.0005	0.0845	6.2321	7.3999	4.1930
UA	0.4817	0.0024	0.0037	2.1654	6.5428	5.9299
LDH	0.9884	0.9582	0.0788	1.2987	1.6571	4.5878
Height	0.8145	0.4240	0.1235	1.6654	1.7714	3.6492
BMI	0.1171	0.0011	0.2146	5.6988	6.8857	2.3070
Gender	0.0049	0.0000	0.1349	6.8988	9.2142	2.6492
Ca	0.0040	0.0000	0.0000	7.0655	9.7142	6.5264
PI	0.0000	0.0001	0.0812	7.7322	8.1285	4.3509
Glu	0.0430	0.0009	0.4154	5.8655	7.1428	2.0614
Apo-A1	0.0036	0.0001	0.4525	7.2655	8.4142	1.8246
Apo-B	0.0013	0.0001	0.6987	7.5822	8.7285	1.4386
DBIL	0.0364	0.0003	0.2629	6.0821	7.6285	2.3070
TC	0.0248	0.0000	0.0382	6.4322	8.9571	5.4387
TBIL	0.1633	0.0323	0.5188	5.4488	5.3857	1.7018
TP	0.3667	0.0946	0.0417	3.0654	2.0428	5.2808

[a] The full forms of all abbreviations are shown in Table 2.

Table 8. Weighting of the 28 features using weighting- and ranking-based hybrid feature selection.

Order	Feature[a]	Weight	Contribution	Cumulative contribution	Weight (0-1)
1	Age	176.31	0.13	0.13	1
2	α-HBD	88.36	0.06	0.19	0.42
3	SCr	83.02	0.06	0.25	0.38
4	LDH	70.59	0.05	0.30	0.30
5	Height	70.32	0.05	0.35	0.30
6	TBIL	66.18	0.05	0.39	0.27
7	CK	59.22	0.04	0.44	0.22
8	Apo-B	55.61	0.04	0.48	0.20
9	CK-MB	54.09	0.04	0.51	0.19
10	Alb	48.60	0.03	0.55	0.15
11	AST	47.49	0.03	0.58	0.15
12	GGP	45.36	0.03	0.61	0.13
13	DBIL	40.76	0.03	0.64	0.10
14	Glu	39.35	0.03	0.67	0.09
15	Gender	37.99	0.03	0.70	0.08
16	ALP	36.26	0.03	0.72	0.07
17	Apo-A1	35.60	0.03	0.75	0.07
18	TP	35.22	0.03	0.77	0.06
19	Ca	34.34	0.02	0.80	0.06
20	TC	34.34	0.02	0.82	0.06
21	BMI	33.90	0.02	0.85	0.06
22	HDL	33.16	0.02	0.87	0.05
23	BUN	32.92	0.02	0.89	0.05
24	PI	32.61	0.02	0.92	0.05
25	TG	32.37	0.02	0.94	0.05
26	UA	30.70	0.02	0.96	0.03
27	LDL	27.46	0.02	0.98	0.01
28	ALT	25.56	0.02	1.00	0

[a] The full forms of all abbreviations are shown in Table 2.

Table 9. Classification performances of support vector machine with different feature selection methods.

Method	Features	Sensitivity (N=398), n (%)	Specificity (N=394), n (%)	Accuracy (N=792), n (%)	Youden index
WRHFS[a]	9	329 (82.7)	317 (80.4)	645 (81.5)	0.63
Information gain	10	297 (74.6)	284 (72.1)	574 (72.5)	0.47
Relief	13	277 (69.6)	290 (73.7)	577 (72.9)	0.43
Standard deviation	20	283 (71.1)	291 (73.9)	580 (73.2)	0.45

[a] WRHFS: weighting- and ranking-based hybrid feature selection.

Table 10. Classification performances of different models with weighting- and ranking-based hybrid feature selection.

Classifier	Sensitivity (N=398), n (%)	Specificity (N=394), n (%)	Accuracy (N=792), n (%)	Youden index
SVM[a]	329 (82.7)	317 (80.4)	645 (81.5)	0.63
Bayes	319 (80.2)	197 (50.02)	520 (65.7)	0.30
CBA[b]	305 (76.6)	300 (76.1)	605 (76.4)	0.53
BPNN[c]	280 (70.4)	220 (55.8)	501 (63.2)	0.26
CART[d]	280 (70.4)	283 (71.8)	562 (71.0)	0.42
C4.5	269 (67.6)	302 (76.6)	571 (72.1)	0.44
ELM[e]	220 (55.3)	249 (63.2)	469 (59.2)	0.19

[a] SVM: support vector machine.

[b] CBA: classification based on associations.

[c] BPNN: back-propagation neural networks.

[d] CART: classification and regression tree.

[e] ELM: extreme learning machine.

We visualized the change trend of the risk levels of ischemic stroke in Figure 2 using the surface fitting technique based on the 9 key features. The synthetic value (SV) indicates the linear combination of the feature value and its weight. The risk of ischemic stroke is reflected in the SV, which is defined as follows:

$$SV=AGE+0.42 \times \alpha\text{-}HBD+0.38 \times SCr+0.3 \times LDH+0.3 \times HEIGHT+0.27 \times TBIL+0.22 \times CK+0.2 \times Apo\text{-}B+0.19 \times CK\text{-}MB$$

where age, α-HBD, and other features indicate the feature values, and 0.42, 0.38, and other values are the weights associated with individual features. Figure 2 presents the surface chart for stroke risk detection, in which the Y axis represents the age between 45 and 90 years, the Z axis represents risk index of suffering from ischemic stroke, and the X axis represents the SV. Figure 3 presents the risk index map for ischemic stroke detection, which is a top view of Figure 2. There were 33 ranks of risk index: "≤1.5" means no risk; ">1.5 but ≤2" means low risk; and ">2" means high risk. Different colors indicate different levels of risks.

Figure 2. A surface chart for risk detection.

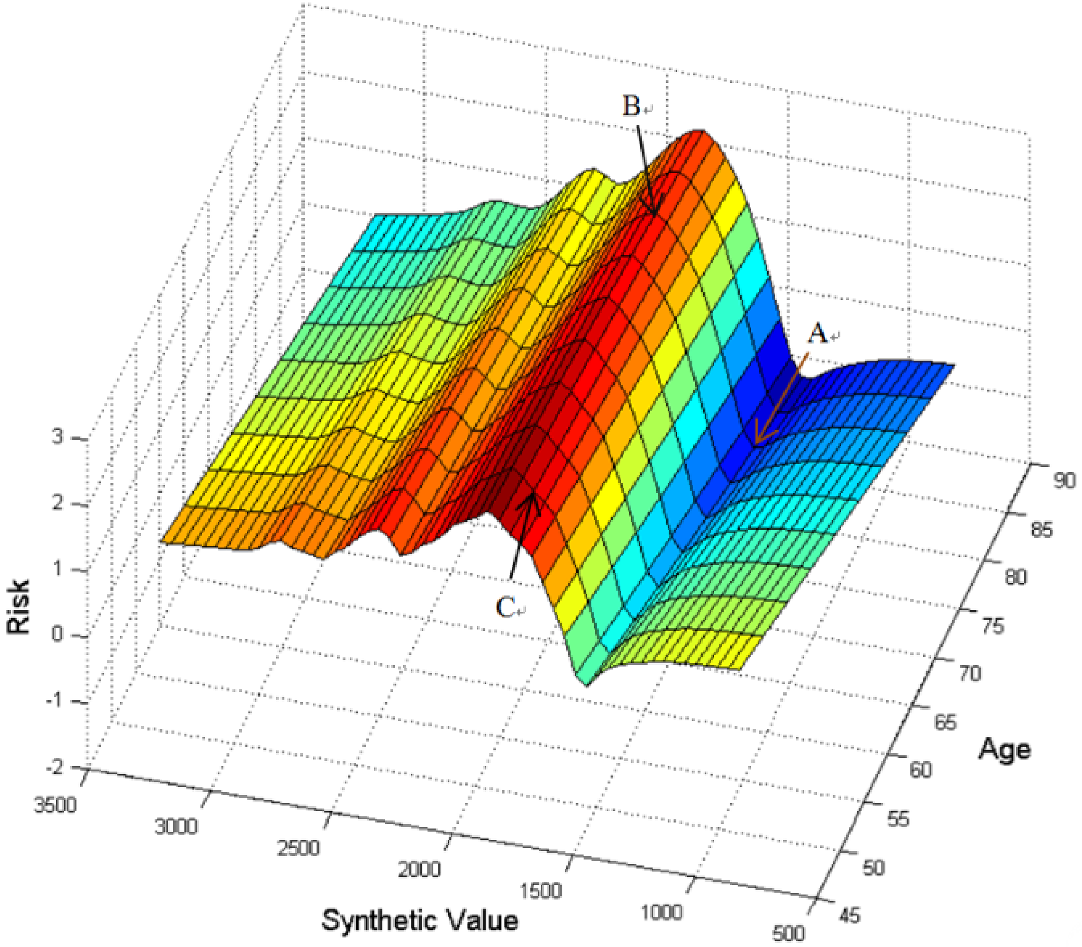
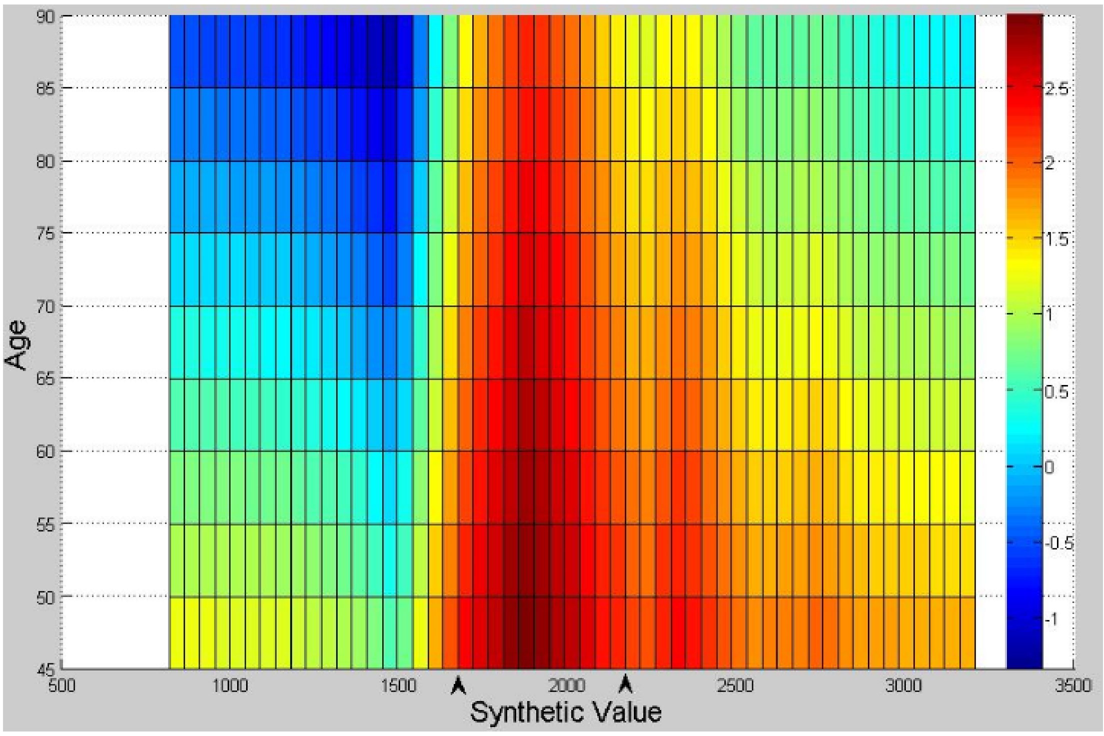


Figure 3. A risk index map for ischemic stroke detection.



Discussion

To address the limitations of existing risk detection models and expensive detection costs in hospitals, we proposed a new feature selection method, namely WRHFS, for risk detection of ischemic stroke. In this study, WRHFS selected features through the guidance of the top 3 filter methods based on the 28 risk factors. It provided an aggregated importance weight for each feature. As shown in Table 8, the top 9 features that achieved sensitivity of 82.7% (329/398) were selected for detecting the risk of ischemic stroke. WRHFS can also evaluate the effectiveness of the existing filter feature selection methods based on effective coefficients and choose the top 3 filter methods. On the basis of the sorted results of the importance weights of individual features, we chose 9 features and produced the change trend of risk levels and the risk index map for ischemic stroke.

Principal Findings

We compared the performance of the proposed feature selection method WRHFS against those of standard deviation, Relief, and information gain. The results revealed that age is the most important influence factor because it has the largest weight value, which is consistent with the literature on stroke [6,13]. Through ranking the features in decreasing order of their importance to the performance of a risk detection model, WHRFS enables us to choose the most effective features that have the highest contributions to the performance measures of a model. Results of evaluation demonstrate that WRHFS can achieve a contribution rate of 0.51 with only the first 9 features, whereas the other 3 traditional feature selection methods require more. As a feature selection method, WRHFS is superior by being able to calculate effectiveness coefficients of individual features.

The contributions of the other 27 risk factors, excluding age, vary in the models constructed by the 4 feature selection methods, including WRHFS, standard deviation, Relief, and information gain. More specifically, the contribution of α-HBD assessed by Relief is significantly greater than that assessed by other feature selection methods, and the contribution of CK was ranked highest by standard deviation but almost 0 by Relief. It implies that a single objective function may not be able to measure the importance of risk factors comprehensively.

Age is the most important feature found in this study. The contribution of age to the model's performance is approximately 13%. The risk of stroke was reflected in the SV. Therefore, age should be integrated in the SV. In general, the risk of ischemic stroke increases with age. As shown in Figure 2, A and B have the same age but are much older than C. However, B has higher risk than A because of the higher SV. In contrast, C is younger than A but has a higher risk than A also because of a higher SV. Therefore, the risk of ischemic stroke is influenced by the SV. A person would have low risk of ischemic stroke if the SV is far from the high-risk interval (HRI; ie, 1675, 2175), which is shown in Figure 3. The findings of this study will not only provide methodological guidance on how to select more effectiveness features for automated detection of stroke risk but also potentially help physicians improve their diagnosis in medical practice.

The major contribution of this research is WRHFS, a new generic feature selection method. WRHFS deploys continuous weighting and ranking of individual features by following the principle of a wrapper approach that integrates the strengths of various filter methods for feature selection. The evaluation shows that WRHFS can result in a superior risk detection model that achieves better performance with fewer features than the existing feature selection methods, demonstrating the effectiveness of WRHFS.

The findings of this study also provided multiple practical implications for physicians. First, the top 9 features are extremely easy to obtain. Physicians can calculate the corresponding SV and easily detect the ischemic stroke risk indexes using the risk index map as an auxiliary diagnostic method. As shown in Figure 3, the range between 1675 and 2175 of the SV (where the black arrow points) can be called the HRI. There seems a parabolic envelope curve. In addition, elderly people whose ages are between 70 and 90 years tend to have a high risk of ischemic stroke, whereas the risk becomes lower when the SV is smaller (800 to 1500) or larger (3000 to 3250). In addition, an automated stroke risk detection platform can be developed easily by use of the above findings for stroke during the physical examination of people.

Limitations

This study has a couple of limitations that offer future research opportunities. First, the acquisition of medical samples is very difficult. We were unable to find data samples that included all of the risk factors that have been discovered in the literature. It would be worthy to conduct a future study with a larger and different dataset with more features to examine if the finding of this research can still hold. Second, we used a straightforward way to aggregate the rankings of individual filter methods, which may or may not be optimal. We plan to explore other means in future research.

Conclusions

Automatic detection of stroke risks has been increasingly studied in recent years. How to select important factors for risk detection models is critical to the model's performance. Existing research on automatic detection of stroke risks through machine learning faces a significant challenge in the selection of effective features as predictive cues. Therefore, how to develop more effective methods for feature selection is critical. This study proposed, developed, and evaluated a new feature selection method, which can help identify the most important features for building effective and parsimonious models for stroke risk detection. The proposed method, WRHFS, provides a novel methodological research contribution and practical implications.

Acknowledgments

This work was supported by the Natural Science Foundation of Shanxi (Grant No 201601D102031) and the National Natural Science Foundation of China (Grant No 6160051296). The authors would like to thank the anonymous reviewers whose comments helped them to greatly improve the final version of the paper.

Authors' Contributions

YZ wrote the full draft of the manuscript and was involved in the execution, concept, and overall responsibility of this study. YZ was involved in the execution and data analyses. DZ revised, reviewed, and approved the final manuscript. WS read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Weighting of the 28 features based on Relief.

[PDF File (Adobe PDF File), 73KB - jmir_v21i4e12437_app1.pdf]

Multimedia Appendix 2

Weighting of the 28 features based on information gain.

[PDF File (Adobe PDF File), 72KB - jmir_v21i4e12437_app2.pdf]

References

1. World Health Organization. Cardiovascular diseases (CVDs) URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2019-02-18] [WebCite Cache ID 76HRz3QR9]
2. Celermajer DS, Chow CK, Marijon E, Anstey NM, Woo KS. Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection. J Am Coll Cardiol 2012 Oct 02;60(14):1207-1216 [FREE Full text] [doi: 10.1016/j.jacc.2012.03.074] [Medline: 22858388]
3. Ning X, Sun J, Jiang R, Lu H, Bai L, Shi M, et al. Increased stroke burdens among the low-income young and middle aged in rural China. Stroke 2017 Dec;48(1):77-83. [doi: 10.1161/STROKEAHA.116.014897] [Medline: 27924051]
4. Acharya UR, Mookiah MR, Vinitha Sree S, Afonso D, Sanches J, Shafique S, et al. Atherosclerotic plaque tissue characterization in 2D ultrasound longitudinal carotid scans for automated classification: a paradigm for stroke risk assessment. Med Biol Eng Comput 2013 May;51(5):513-523. [doi: 10.1007/s11517-012-1019-0] [Medline: 23292291]
5. Papadopoulos H, Kyriacou E, Nicolaides A. Unbiased confidence measures for stroke risk estimation based on ultrasound carotid image analysis. Neural Comput Applic 2016 Sep 24;28(6):1209-1223. [doi: 10.1007/s00521-016-2590-3]
6. Zhang Y, Song W, Li S, Fu L, Li S. Risk detection of stroke using a feature selection and classification method. IEEE Access 2018;6:31899-31907. [doi: 10.1109/ACCESS.2018.2833442]
7. Cho S, Ku J, Cho YK, Kim IY, Kang YJ, Jang DP, et al. Development of virtual reality proprioceptive rehabilitation system for stroke patients. Comput Methods Programs Biomed 2014;113(1):258-265. [doi: 10.1016/j.cmpb.2013.09.006] [Medline: 24183070]
8. Saba L, Banchhor SK, Londhe N, Araki T, Laird J, Gupta A, et al. Web-based accurate measurements of carotid lumen diameter and stenosis severity: an ultrasound-based clinical tool for stroke risk assessment during multicenter clinical trials. Comput Biol Med 2017 Dec 01;91:306-317. [doi: 10.1016/j.compbiomed.2017.10.022] [Medline: 29107894]
9. Park E, Chang HJ, Nam HS. Use of machine learning classifiers and sensor data to detect neurological deficit in stroke patients. J Med Internet Res 2017 Dec 18;19(4):e120 [FREE Full text] [doi: 10.2196/jmir.7092] [Medline: 28420599]
10. Ab Malik N, Mohamad Yatim S, Lam OL, Jin L, McGrath CP. Effectiveness of a web-based health education program to promote oral hygiene care among stroke survivors: randomized controlled trial. J Med Internet Res 2017 Dec 31;19(3):e87 [FREE Full text] [doi: 10.2196/jmir.7024] [Medline: 28363880]
11. O'Brien MK, Shawen N, Mummidisetty CK, Kaur S, Bo X, Poellabauer C, et al. Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting. J Med Internet Res 2017 Dec 25;19(5):e184 [FREE Full text] [doi: 10.2196/jmir.7385] [Medline: 28546137]
12. Breccia M, Molica M, Zacheo I, Serrao A, Alimena G. Application of systematic coronary risk evaluation chart to identify chronic myeloid leukemia patients at risk of cardiovascular diseases during nilotinib treatment. Ann Hematol 2015 Mar;94(3):393-397. [doi: 10.1007/s00277-014-2231-9] [Medline: 25304102]
13. Tunstall-Pedoe H. Cardiovascular risk and risk scores: ASSIGN, Framingham, QRISK and others: how to choose. Heart 2011;97(6):442-444. [Medline: 21339319]

14. Hu T, Yao L, Reynolds K, Whelton PK, Niu TH, Li SX, et al. The effects of a low-carbohydrate diet vs a low-fat diet on novel cardiovascular risk factors: a randomized controlled trial. Nutrients 2015 Sep 17;7(9):7978-7994 [FREE Full text] [doi: 10.3390/nu7095377] [Medline: 26393645]
15. Pencina M, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation 2009 Jun 23;119(24):3078-3084 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.816694] [Medline: 19506114]
16. Flueckiger P, Longstreth W, Herrington D, Yeboah J. Revised framingham stroke risk score, nontraditional risk markers, and incident stroke in a multiethnic cohort. Stroke 2018 Dec;49(2):363-369 [FREE Full text] [doi: 10.1161/STROKEAHA.117.018928] [Medline: 29311270]
17. Yang X, Li S, Zhao X, Liu L, Jiang Y, Li Z, et al. Atrial fibrillation is not uncommon among patients with ischemic stroke and transient ischemic stroke in China. BMC Neurol 2017 Dec 04;17(1):207 [FREE Full text] [doi: 10.1186/s12883-017-0987-y] [Medline: 29202727]
18. Chang CS, Su SL, Kuo CL, Huang CS, Tseng WM, Lin S, et al. Cyclophilin A: a predictive biomarker of carotid stenosis in cerebral ischemic stroke. Curr Neurovasc Res 2018;15(2):111-119. [doi: 10.2174/1567202615666180516120959] [Medline: 29766804]
19. Zhang X, Attia J, D'Este C, Yu X, Wu X. A risk score predicted coronary heart disease and stroke in a Chinese cohort. J Clin Epidemiol 2005 Sep;58(9):951-958. [doi: 10.1016/j.jclinepi.2005.01.013] [Medline: 16085199]
20. Di Iorio B, Di Micco L, Torraca S, Sirico M, Guastaferro P, Chiuchiolo L, et al. Variability of blood pressure in dialysis patients: a new marker of cardiovascular risk. J Nephrol 2013;26(1):173-182. [doi: 10.5301/jn.5000108] [Medline: 22419232]
21. Hadaegh F, Mohebi R, Khalili D, Hasheminia M, Sheikholeslami F, Azizi F. High normal blood pressure is an independent risk factor for cardiovascular disease among middle-aged but not in elderly populations: 9-year results of a population-based study. J Hum Hypertens 2013 Jan;27(1):18-23. [doi: 10.1038/jhh.2011.112] [Medline: 22217674]
22. Deng L, Dai H. Epidemiology and risk factors of cardiovascular disease in Chuxiong city. Chinese journal of cardiovascular research 2017;15(7):610-614.
23. Wang L, Liu J, Yang Y. Summary of stroke prevention report of China in 2017. Chin J Cerebrovasc Dis 2018;15(11):611-617.
24. Wang L, Wang J, Pen B. Summary of stroke prevention report of China in 2016. Chin J Cerebrovasc Dis 2017;14(4):217-224 [FREE Full text]
25. Chen W, Gao R, Liu L. Summary of cardiovascular disease report of China in 2017. Chinese Circulation Journal 2019 Jan;99:1-26.
26. Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, USA-PRC Collaborative Study of CardiovascularCardiopulmonary Epidemiology Research Group, China Multicenter Collaborative Study of Cardiovascular Epidemiology Research Group. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation 2006 Nov 21;114(21):2217-2225. [doi: 10.1161/CIRCULATIONAHA.105.607499] [Medline: 17088464]
27. PRC-USA. An epidemiological study of cardiovascular and cardiopulmonary disease risk factors in four populations in the People's Republic of China. Baseline report from the P.R.C.-U.S.A. Collaborative Study. People's Republic of China--United States Cardiovascular and Cardiopulmonary Epidemiology Research Group. Circulation 1992 Mar;85(3):1083-1096. [Medline: 1537106]
28. Li M, Hou W, Zhang X, Hu L, Tang Z. Hyperuricemia and risk of stroke: a systematic review and meta-analysis of prospective studies. Atherosclerosis 2014 Feb;232(2):265-270. [doi: 10.1016/j.atherosclerosis.2013.11.051] [Medline: 24468137]
29. Chao TF, Liu CJ, Chen SJ, Wang KL, Lin YJ, Chang SL, et al. Hyperuricemia and the risk of ischemic stroke in patients with atrial fibrillation--could it refine clinical risk stratification in AF? Int J Cardiol 2014 Jan 01;170(3):344-349. [doi: 10.1016/j.ijcard.2013.11.011] [Medline: 24290426]
30. Lloyd-Jones DM, Goff DC, Stone NJ. Guidelines for cardiovascular risk assessment and cholesterol treatment. J Am Med Assoc 2014 Jun 04;311(21):2235. [doi: 10.1001/jama.2014.4072] [Medline: 24893098]
31. Wen C, David Cheng TY, Tsai S, Chan H, Hsu H, Hsu C, et al. Are Asians at greater mortality risks for being overweight than Caucasians? Redefining obesity for Asians. Public Health Nutr 2009 Apr;12(4):497-506. [doi: 10.1017/S1368980008002802] [Medline: 18547457]
32. Ali AT, Chirambo G, Penny C, Paiker JE, Ikram F, Psaras G, et al. Ethnic differences in pre-adipocyte intracellular lipid accumulation and alkaline phosphatase activity. Clin Chim Acta 2015 Jan 01;438:382-387. [doi: 10.1016/j.cca.2014.09.022] [Medline: 25281857]
33. Yang Y, Li J, Chen JC, Ca J, Lu XF, Chen SF, et al. Effect of elevated total cholesterol level and hypertension on the risk of fatal cardiovascular disease: a cohort study of Chinese steelworkers. Chin Med J (Engl) 2011 Nov;124(22):3702-3706. [Medline: 22340227]
34. Guo J, Li W, Wang Y, Chen T, Teo K, Liu L, INTERHEART China study Investigators. Influence of socioeconomic status on acute myocardial infarction in the Chinese population: the INTERHEART China study. Chin Med J (Engl) 2012 Dec;125(23):4214-4220. [Medline: 23217389]
35. Xu Z, King I, Lyu M, Jin R. Discriminative semi-supervised feature selection via manifold regularization. IEEE Trans Neural Netw 2010 Jul;21(7):1033-1047. [doi: 10.1109/TNN.2010.2047114] [Medline: 20570772]

36. Zhao Z, Wang L, Liu H, Ye J. On similarity preserving feature selection. IEEE Trans Knowl Data Eng 2013 Mar;25(3):619-632. [doi: 10.1109/TKDE.2011.222]
37. Yilmaz Eroglu D, Kilic K. A novel hybrid genetic local search algorithm for feature selection and weighting with an application in strategic decision making in innovation management. Inf Sci 2017 Sep;405:18-32. [doi: 10.1016/j.ins.2017.04.009]
38. Doquire G, Verleysen M. A graph Laplacian based approach to semi-supervised feature selection for regression problems. Neurocomputing 2013 Dec;121(18):5-13. [doi: 10.1016/j.neucom.2012.10.028]
39. Alalga A, Benabdeslem K, Taleb N. Soft-constrained Laplacian score for semi-supervised multi-label feature selection. Knowl Inf Syst 2015 May 17;47(1):75-98. [doi: 10.1007/s10115-015-0841-8]
40. Jiang W, Chung F. A trace ratio maximization approach to multiple kernel-based dimensionality reduction. Neural Netw 2014 Jan;49(1):96-106. [doi: 10.1016/j.neunet.2013.09.004] [Medline: 24211342]
41. Zhu P, Xu Q, Hu Q, Zhang C. Co-regularized unsupervised feature selection. Neurocomputing 2018 Jan;275:2855-2863. [doi: 10.1016/j.neucom.2017.11.061]
42. Sadri A, Ren Y, Salim F. Information gain-based metric for recognizing transitions in human activities. Pervasive Mob Comput 2017 Jul;38:92-109. [doi: 10.1016/j.pmcj.2017.01.003]
43. Sakar C, Kursun O, Gurgen F. A feature selection method based on kernel canonical correlation analysis and the minimum Redundancy–Maximum Relevance filter method. Expert Syst Appl 2012 Feb;39(3):3432-3437. [doi: 10.1016/j.eswa.2011.09.031]
44. Huang X, Zhang L, Wang B, Zhang Z, Li F. Feature weight estimation based on dynamic representation and neighbor sparse reconstruction. Pattern Recognition 2018;81:388-403. [doi: 10.1016/j.patcog.2018.03.014]
45. Chyzhyk D, Savio A, Graña M. Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI. Neurocomputing 2014 Mar;128:73-80. [doi: 10.1016/j.neucom.2013.01.065]
46. Hancer E, Xue B, Zhang M. Differential evolution for filter feature selection based on information theory and feature ranking. Knowl Based Syst 2018 Jan;140:103-119. [doi: 10.1016/j.knosys.2017.10.028]
47. Sun Y, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans Pattern Anal Mach Intell 2010 Sep;32(9):1610-1626 [FREE Full text] [doi: 10.1109/TPAMI.2009.190] [Medline: 20634556]
48. Yang J, Liu YL, Feng CS, Zhu GQ. Applying the Fisher score to identify Alzheimer's disease-related genes. Genet Mol Res 2016 Jun 27;15(2):- [FREE Full text] [doi: 10.4238/gmr.15028798] [Medline: 27420981]
49. Farshad M, Sadeh J. Transmission line fault location using hybrid wavelet-Prony method and relief algorithm. International Journal of Electrical Power & Energy Systems 2014 Oct;61(1):127-136. [doi: 10.1016/j.ijepes.2014.03.045]
50. Kononenko I. Estimating attributes: analysisextensions of RELIEF. In: Francesco B, Luc DR, editors. European Conference on Machine Learning. Berlin Heidelberg: Springer Press; 1994:171-182.
51. Huan L, Lei Y. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans Knowl Data Eng 2005 Apr;17(4):491-502. [doi: 10.1109/TKDE.2005.66]
52. Lin Y, Hu Q, Liu J, Li J, Wu X. Streaming feature selection for multilabel learning based on fuzzy mutual information. IEEE Trans Fuzzy Syst 2017 Dec;25(6):1491-1507. [doi: 10.1109/TFUZZ.2017.2735947]
53. Uysal A, Gunal S. A novel probabilistic feature selection method for text classification. Knowl Based Syst 2012 Dec;36(6):226-235. [doi: 10.1016/j.knosys.2012.06.005]
54. Wang Y, Feng L. Hybrid feature selection using component co-occurrence based feature relevance measurement. Expert Systems with Applications 2018 Jul;102:83-99. [doi: 10.1016/j.eswa.2018.01.041]
55. Lu H, Chen J, Yan K, Jin Q, Xue Y, Gao Z. A hybrid feature selection algorithm for gene expression data classification. Neurocomputing 2017 Sep;256:56-62. [doi: 10.1016/j.neucom.2016.07.080]
56. Azizi R, Attaran B, Hajnayeb A, Ghanbarzadeh A, Changizian M. Improving accuracy of cavitation severity detection in centrifugal pumps using a hybrid feature selection technique. Measurement 2017 Oct;108:9-17. [doi: 10.1016/j.measurement.2017.05.020]
57. Zhang Y, Liu W, Ge S, Yan M, Wang S, Yu J, et al. Multiplexed sandwich immunoassays using flow-injection electrochemiluminescence with designed substrate spatial-resolved technique for detection of tumor markers. Biosens Bioelectron 2013 Mar 15;41:684-690. [doi: 10.1016/j.bios.2012.09.044] [Medline: 23062558]
58. Awan Z, Shoaib A, Khan K. Variations in total phenolics and antioxidant enzymes cause phenotypic variability and differential resistant response in tomato genotypes against early blight disease. Scientia Horticulturae 2018 Sep;239:216-223. [doi: 10.1016/j.scienta.2018.05.044]
59. Pagliano P, Boccia G, De Caro F, Esposito S. Bacterial meningitis complicating the course of liver cirrhosis. Infection 2017 Dec;45(6):795-800. [doi: 10.1007/s15010-017-1039-7] [Medline: 28616745]
60. Bachlechner U, Boeing H, Haftenberger M, Schienkiewitz A, Scheidt-Nave C, Vogt S, et al. Predicting risk of substantial weight gain in German adults-a multi-center cohort approach. Eur J Public Health 2017 Dec 01;27(4):768-774 [FREE Full text] [doi: 10.1093/eurpub/ckw216] [Medline: 28013243]
61. Stokes ME, Visweswaran S. Application of a spatially-weighted Relief algorithm for ranking genetic predictors of disease. BioData Min 2012 Dec 03;5(1):20 [FREE Full text] [doi: 10.1186/1756-0381-5-20] [Medline: 23198930]

62. Busing FM, Weaver B, Dubois S. 2 × 2 Tables: a note on Campbell's recommendation. Stat Med 2016 Apr 15;35(8):1354-1358. [doi: 10.1002/sim.6808] [Medline: 26576745]
63. Thompson DR, Martin CR. Bayes' theorem and its application to cardiovascular nursing. Eur J Cardiovasc Nurs 2017 Dec;16(8):659-661. [doi: 10.1177/1474515117712317] [Medline: 28537096]
64. Wang L, Haug PJ, Del Fiol G. Using classification models for the generation of disease-specific medications from biomedical literature and clinical data repository. J Biomed Inform 2017 Dec;69:259-266 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.014] [Medline: 28435015]
65. Kuznetsov V, Yaroslavskaya E, Krinochkin D, Teffenberg D, Dyachkov S. Artificial neural network for prediction of coronary artery disease in patients with abnormal myocardial perfusion. Eur J Heart Failure 2018;20:484-484 [illegible]
66. Henrard S, Speybroeck N, Hermans C. Classification and regression tree analysis vs multivariable linear and logistic regression methods as statistical tools for studying haemophilia. Haemophilia 2015 Nov;21(6):715-722. [doi: 10.1111/hae.12778] [Medline: 26248714]
67. Jothikumar R. C4.5 classification algorithm with back-track pruning for accurate prediction of heart disease. Biomed Res 2016;27:107-111 [FREE Full text]
68. Liu T, Hu L, Ma C, Wang Z, Chen H. A fast approach for detection of erythemato-squamous diseases based on extreme learning machine with maximum relevance minimum redundancy feature selection. Int J Syst Sci 2013 May 28;46(5):919-931. [doi: 10.1080/00207721.2013.801096]

Abbreviations

BMI: body mass index
BPNN: back-propagation neural networks
CK: creatine kinase
CVD: cardiovascular disease
HDL: high-density lipoprotein
HRI: high-risk interval
SVM: support vector machine
SV: synthetic value
TC: total cholesterol
WRHFS: weighting- and ranking-based hybrid feature selection

Edited by CL Parra-Calderón; submitted 07.10.18; peer-reviewed by Z Yan, X Liu, I Kagashe; comments to author 01.12.18; revised version received 04.01.19; accepted 26.01.19; published 02.04.19.

Please cite as:
Zhang Y, Zhou Y, Zhang D, Song W
A Stroke Risk Detection: Improving Hybrid Feature Selection Method
J Med Internet Res 2019;21(4):e12437
URL: https://www.jmir.org/2019/4/e12437/
doi:10.2196/12437
PMID:30938684

Original Paper

Use of Rideshare Services to Increase Participant Recruitment and Retention in Research: Participant Perspectives

Eleanor Ladd Schneider Leavens[1,2], MS; Elise Marie Stevens[1], PhD; Emma Irene Brett[2], MS; Neil Molina[1], BS; Thad Ryan Leffingwell[2], PhD; Theodore Lee Wagener[1,3], PhD

[1]Oklahoma Tobacco Research Center, Oklahoma City, OK, United States
[2]Oklahoma State University, Department of Psychology, Stillwater, OK, United States
[3]University of Oklahoma Health Sciences Center, Department of Pediatrics, Oklahoma City, OK, United States

Corresponding Author:
Eleanor Ladd Schneider Leavens, MS
Oklahoma Tobacco Research Center
655 Research Parkway
Oklahoma City, OK,
United States
Phone: 1 9139728841
Email: leavens@okstate.edu

Abstract

Background: Recruitment and retention of participants are important factors in empirical studies. Methods that increase recruitment and retention can reduce costs and burden on researchers related to the need for over-recruitment because of attrition. Rideshare services such as Uber and Lyft are a potential means for decreasing this burden.

Objective: This study aimed to understand the role rideshare utilization plays in participant recruitment and retention in research trials.

Methods: Data are presented for a study (*N*=42) in which rideshare services were utilized for participant transportation to and from study visits during a 2-session, in-laboratory research study.

Results: Retention at visit 2 was greater than 95% (42/44) in the initial study. In a follow-up survey of the participants from the original trial, participants (N=32) reported that the rideshare service was an important reason they returned for all study visits. Participants reported whether they would prefer differing levels of additional monetary compensation or a ride from a rideshare service. When the additional compensation was less than US $15, participants reported a preference for the rideshare service.

Conclusions: Rideshare services may represent a relatively low cost means for increasing study retention. Specifically, findings indicate that rideshare services may not be crucial for initial participant recruitment but for their retention in multi-visit studies.

(J Med Internet Res 2019;21(4):e11166) doi:10.2196/11166

KEYWORDS

rideshare service; recruitment; retention; attrition; transportation

Introduction

Background

In health, behavioral, and social sciences, human laboratory and randomized controlled trials are imperative to further science and interventions [1-4]. Two critical aspects of research studies are recruitment and retention [5]. Recruitment is the process by which potential research participants are made aware of and then enrolled in the study [5], whereas retention refers to participants staying in the study and completing study visits [6]. When researchers fail to recruit and retain participants, findings can be invalid, inconclusive, and insufficient to answer research questions [7-9]. In addition, attrition can be costly and result in using greater resources, extending studies, and in some cases, terminating studies prematurely [10].

Research suggests that telephone reminders and financial incentives are advantageous ways to recruit and retain participants [11]. Specifically, studies have shown that participants increase their willingness to participate when compensation increases [12,13], regardless of the risk of adverse events that may result from study participation [12]. Recruitment and retention strategies that reduce participant burden, including costs associated with transportation, are likely to impact desirability of study participation.

It is also likely that the content area impacts recruitment and retention of participants. On one hand, because of the unique challenges associated with retaining participants in substance abuse studies, the threshold for good retention in these studies has been set relatively low (70%) [14]. However, in studies utilizing alcohol and tobacco administration, retention rates are relatively high (90%-96%) [15,16]. However, these referenced studies also required participants to complete baseline screening visits and did not include participants who did not return following screening, which likely increased retention rates.

A recent study proposed a conceptual framework specific to improving recruitment and retention in tobacco and alcohol research, including (1) creating a team mindset that promotes regular and positive communication with participants, (2) leveraging technology, and (3) increasing efforts to contact nonresponsive participants [17]. Applying such techniques as applicable to a specific sample may improve retention by establishing an alliance between the study team and the participant and reducing participant burden. In addition, literature on retention suggests that using more strategies across several separate categories (eg, visit characteristics, study personnel, and nonfinancial incentives) will result in improved retention [18].

With the advent, popularity, and low cost of rideshare services such as Uber and Lyft, it is important to understand if and how such services may represent a novel and advantageous strategy to recruit and retain participants. Past studies among low-income urban participants have shown that reimbursing for taxi services was not always an effective method for recruitment and retention because of unreliability of taxis [8]. However, rideshare services may be more advantageous given the researcher’s capability to order the rides remotely at the scheduled time, track participants’ rides to the study location, and set up an account to facilitate hassle-free payment for the service. Use of rideshare services may also increase safety for studies in which acute intoxication is necessary and driving would put participants in harm’s way. Finally, use of rideshare services reduces participants’ burden and is consistent with the suggestions within Smith and colleagues’ framework for increasing retention [17].

Objectives

The aim of this study was to assess participants’ perceptions of the use of a rideshare service in terms of the impact it had on decisions to return to study visits in a recently completed, multi-visit study. We also examined how providing rideshare services in future studies would influence participants’ decisions to participate. Finally, we aimed to understand whether differing levels of additional compensation or rideshare services would be better for recruitment and retention.

Methods

Participants and Procedure

This study recruited participants from a recently completed study [19]. The purpose of the original study was to understand the impact of acute alcohol intoxication on waterpipe smoking patterns and toxicant exposure. The completed research study recruited 21 dyads (*N*=42) of current waterpipe smokers and drinkers for a 2-session, in-laboratory study. Each visit included survey completion, 2 blood draws, breath tests (breath alcohol concentration and carbon monoxide), and alcohol or placebo beverage administration, followed by a waterpipe smoking session lasting up to 2 hours. Of the 44 participants, 42 (95.5%) were retained in the study. Retention methods included regular calls to participants, relationship building between research staff and participants, fair compensation (US $125 per visit) with a bonus (US $20) for completing both study visits, and transportation to and from study visits via a rideshare service.

Participants who completed both study visits in the original study were invited to provide feedback on their experiences, with the primary aim of understanding the role of the provision of a rideshare service in their choice to complete both study visits. Before completing study procedures, participants provided informed consent. All data were collected remotely via a brief, Web-based survey. Participants were compensated with a US $5 gift card. Of the 42 participants who completed the original study, 32 (mean age 25.7, SD 3.0; 58% male; 79% white) completed this study. The university’s institutional review board approved all study procedures.

Use of Rideshare

All participants were required to utilize rideshare services (eg, Uber and Lyft) for their transportation to and from the research site. This requirement not only facilitated the double-blind nature of the study but also enhanced participant safety following alcohol administration. An Uber account was created for the purpose of this research, and all rides were placed from this account. Participants were required to be picked up from and dropped off at their home address to limit the likelihood participants would drive following participation. Pick up from work or other locations was not allowed because of the likelihood that participants would be dropped off at their car following study participation. Ten to 20 min before the scheduled study visit, the research staff contacted participants via phone to ensure they were ready for the ride request to be placed. If confirmed, research staff placed a request for a rideshare service to pick up the participant and bring them to the laboratory. As participants confirmed they were available for pick up, only 2 rides necessitated cancelling during the study. One cancellation was due to the participant having difficulty locating the vehicle, and the other was due to the participant necessitating additional time. Both rides were able to be rescheduled soon after the original ride was declined. Following each visit, research staff placed a request for the rideshare service to pick up the participants at the laboratory and take them to their home addresses.

Table 1. Importance of recruitment and retention strategies compared with provision of rideshare services (N= 32).

Variables	Recruitment and retention strategies		
	Mean (SD)	*t* test	*P* value
Reasons for study completion			
Rideshare service was provided[a]	5.75 (1.70)	—[b]	—
The study visits were in the evening	5.47 (1.90)	0.64	.53
The staff was nice	6.47 (1.30)	−2.26	.03
I received reminder short message service (SMS) text messages	4.97 (1.84)	2.60	.01
I received reminder calls from staff	4.66 (1.81)	2.86	.01
Alcohol was provided at study visits	4.56 (2.23)	2.65	.01
Hookah was provided at study visits	5.28 (1.42)	1.17	.25
The compensation was fair	6.28 (0.96)	−1.78	.08
I would feel bad if I did not attend all visits	5.72 (1.69)	0.07	.94
I got to complete study with my friend	6.31 (0.90)	−1.74	.09
The study visits were fun	6.47 (0.88)	−2.35	.03
Intentions for future study participation			
Rideshare service was provided[a]	4.13 (0.75)	—	—
The study visits were in the evening	4.38 (0.71)	−1.61	.12
The staff was nice	4.56 (0.62)	−2.95	.01
I received reminder SMS text messages	4.06 (0.67)	0.44	.66
I received reminder calls from staff	3.88 (0.79)	1.54	.13
Alcohol was provided at study visits	4.25 (0.84)	−0.89	.38
Hookah was provided at study visits	4.13 (0.79)	0.00	>.99
The compensation was fair	4.81 (0.40)	−4.98	<.01
I got to complete study with my friend	4.63 (0.55)	−3.22	.003
The study visits were fun	4.69 (0.54)	−3.97	<.001

[a] Reasons for study completion were measured on a 1 (*not at all important*) to 7 (*extremely important*) scale. Intention for future study participation were measured on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.

[b] All *t* tests compared use or rideshare to each other strategy, therefore no values are included for the rideshare test as there is no comparator.

Confidentiality Considerations

Participants consented to the use of the rideshare service at screening and via the main consent form. Specifically, participants were informed a rideshare service would be utilized for transportation but that only their first name and address would be provided to the rideshare service. Drivers were not informed that participants were enrolled in a research study. Participants were further informed that the researcher would place the ride requests but that their participation remained voluntary and could be discontinued at any time and with no penalty to participants.

Measures

Reasons for Study Completion

Participants completed 11 items assessing the importance of different recruitment and retention strategies in their decision to complete both visits of the original study. Items were rated on a Likert scale from 1 (*not at all important*) to 7 (*extremely important*). See Table 1 for a complete list of recruitment and retention strategies that were assessed.

Intention for Future Study Participation

Intention for future study participation was measured by 10 items. Participants reported their agreement with each item. Each item completed the sentence beginning with “I would participate in another study like this if...” Response options ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). See Table 1 for a complete list of recruitment and retention strategies that were assessed.

Multiple-Choice Procedure

To further understand participant preferences for rideshare services in research, participants completed a multiple-choice procedure (MCP) [20] task in which they were presented with the choice between varying levels of additional compensation or a ride to and from the study via a rideshare service. The purpose of this item was to assess participants’ interest in the rideshare service versus additional compensation in some future study. The prompt did not specify the amount of compensation for the hypothetical future study, only the additional compensation amounts versus rideshare provision. Participants were instructed: " *Imagine you have been invited to participate in an in-person research study similar to the one you previously completed in our laboratory. Below is a list of monetary values and free Uber rides. Please choose between the monetary value and Uber ride for each set. In other words, for each set, would you rather have the money or a free Uber ride to your study visit?"* Monetary values ranged from US $0.00/free to US $1000. The crossover value, or point where a switch in preference occurred from the rideshare service to the monetary value, was used to indicate the importance of rideshare compared with additional compensation.

Data Analysis

For outcomes including reasons for returning to study visits and intentions for future study participants, means and SDs were calculated for each item. The mean for the rideshare item was compared with the means for all other recruitment and retention strategy items using paired samples t tests. Significance was set to $P<.05$. For the MCP, descriptive information regarding the frequency of preferences for additional compensation or the rideshare service is presented.

Results

Reasons for Study Completion

Participants reported that the provision of transportation via a rideshare service was an important reason they returned for all study visits (mean 5.75, SD 1.70). Participants reported that the provision of a rideshare service was more important in their decision to complete all visits than reminder short message service (SMS) text messages from staff (mean 4.97, SD 1.84; t_{31}=2.60; P=.01), reminder calls from staff (mean 4.66, SD 1.81; t_{31}=2.86; P=.007), and alcohol being provided at study visits (mean 4.56, SD 2.23; t_{31}=2.65; P=.01). However, compared with the provision of a rideshare service, participants rated the staff being nice (mean 6.47, SD 1.30; t_{31}=−2.26; P=.03) and the visits being fun (mean 6.47, SD 0.88; t_{31}=−2.35; P=.03) as more important in their completion of all study visits.

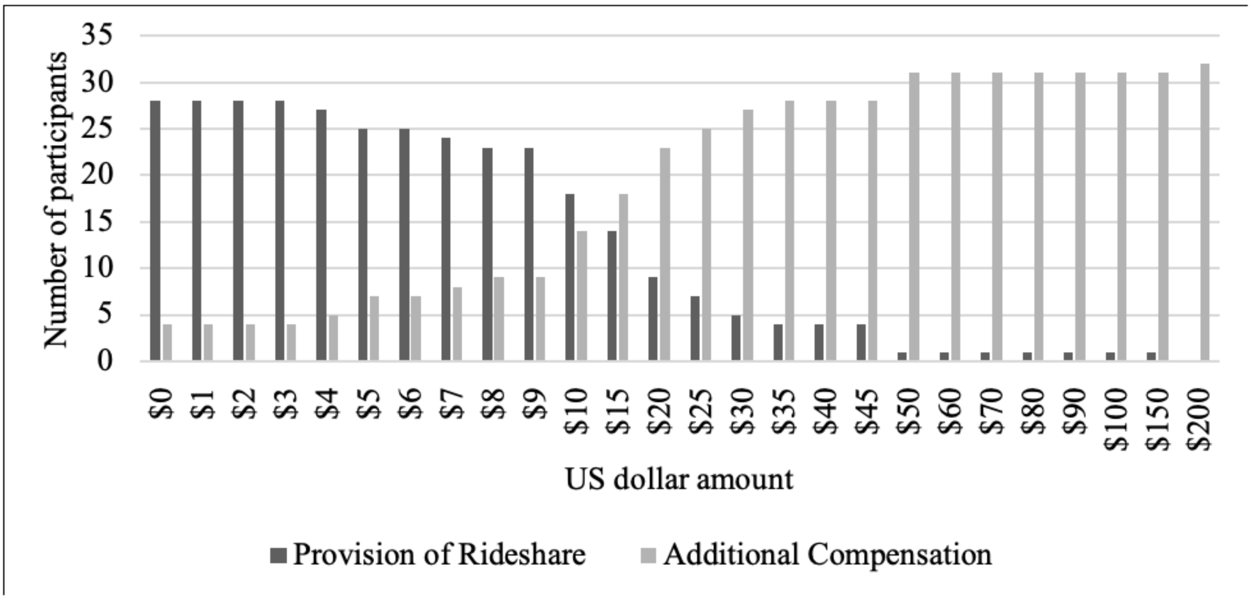
Intention for Future Study Participation

Overall, participants reported that they would participate in a similar study that offered rideshare services in the future (mean 4.13, SD 0.75). However, compared with other recruitment strategies, participants reported a preference for nice staff (mean 4.56, SD 0.62; t_{31}=−2.95; P=.006), fair compensation (mean 4.81, SD 0.40; t_{31}=4.98; $P \leq .001$), the option to participate with a friend (mean 4.63, SD 0.55; t_{31}=3.22; P=.003), and fun study visits (mean 4.69, SD 0.54; t_{31}=−3.97; $P \leq .001$) compared with the provision of rideshare services to and from study visits. See Table 1 for complete results.

Multiple-Choice Procedure

The crossover point on the MCP was observed from US $10 to US $15 such that at levels of additional compensation below US $15, participants showed a preference for the rideshare service. However, participants showed a preference for compensation when the monetary value exceeded US $15 (see Figure 1).

Figure 1. Multiple Choice Procedure—Rideshare versus compensation crossover.



Discussion

Principal Findings

This study is the first to examine participants' perceptions of the use of a rideshare service on their decision to return to study visits and participate in future studies. Recruitment and retention are 2 highly important aspects of successful research [5], and rideshare services may represent a new way to increase participant engagement. In this study, provision of rideshare services was reported to be an important reason participants completed all visits of the original study. Participants also reported they would be interested in completing a future study that provided transportation via a rideshare service. Provision of rideshare services was rated as more important for continued participation than other common recruitment and retention strategies, such as providing reminders for study visits. Alternatively, when considering strategies that would be important in their decision to participate in future studies, participants rated fair compensation and an enjoyable study visit as more important than provision of rideshare services. These seemingly discrepant findings may indicate that rideshare services may not be crucial for the initial recruitment of participants into studies, but for their retention in studies that require more than 1 on-site visit.

Past research has shown that provision of taxi services is not an effective method for recruitment or retention because of the unreliability of taxis [8], but participants in this study reported provision of rideshare was an important reason they returned for study visits, indicating rideshare may be preferred. This difference may be because of reduced participant burden because the researcher is facilitating provision of the ride. In addition, as app-based rideshare services continue to increase in popularity, more people will have experience using them, and thus, be more comfortable with their utilization in research studies. It is currently unknown how perceptions of rideshare versus taxi services may or may not account for these discrepant findings. The use of rideshare services may be a cost-effective way to retain participants, particularly given the high costs both to the researcher and the integrity of the study associated with attrition [7-9]. We investigated the trade-off between providing additional compensation and providing transportation via a rideshare service. The crossover point may indicate that additional compensation is more beneficial than provision of rideshare services at values greater than US $15 but that rideshare services may be more effective if participants live close to the study site and rides cost less than US $10 per participant.

The results of this study, coupled with the outstanding retention rate (95%; 42/44) in the original study, suggest that provision of transportation via rideshare services may be a means for increasing retention that should be shared with other research teams. The potential decreased costs and burdens on research staff related to a decreased need to over-recruit to address attrition may result in significant saved costs. Specifically, saved costs would include those associated with recruiting and compensating additional participants because of decreased attrition. Furthermore, compared with taxis, rideshare services are often less expensive, resulting in additional saved costs. Research staff can also be aware of exact arrival time of participants, given that the rideshare services provide real-time locations of the transportation. In addition, it is likely that the avoided costs of over-recruiting to replace participants lost to follow-up outweigh the costs associated with providing rideshare services. This strategy may be particularly helpful for recruiting and retaining individuals with inconsistent methods of transportation or financial barriers that would make obtaining reliable transportation and attending study visits difficult. Furthermore, utilization of rideshare services in addiction studies in which substance administration is required can reduce additional time and resource burden on researchers. Rideshare services represent a means by which to ensure participants arrive home safely. In the case of alcohol administration studies, particularly those using low alcohol doses, use of rideshare services may limit the need for research staff to remain in the lab with participants until their breath alcohol concentration is 0.000.

Limitations

Although this study is an important step in understanding the integration and use of rideshare services in research, this study has 2 primary limitations. First, we did not utilize a control and are therefore unable to compare differences in recruitment and retention in studies that did and did not use rideshare services. However, the study demonstrated exceptional retention relative to typical studies in the literature, indicating the retention procedures were successful. It is impossible to conclude with strong inference that rideshare was the critical ingredient, but participant reports are consistent with this conclusion. Second, this study may not generalize to other study designs, studies with different aims and methods, or more highly diverse participant populations. Third, although approximately 76% (32/42) of eligible participants participated in this study, we were not able to make conclusions regarding those who chose not to participate in this study and similarly did not assess reasons for discontinuing participation in the original trial.

Conclusions

Despite the large number of studies that require multiple in-lab visits, there has been little research on novel retention and recruitment strategies, an area that is critical for the success of such research. As illustrated in this study, use of rideshare services for in-lab studies may be a worthwhile strategy for increasing retention in research. Utilization of rideshare services should be considered to supplement existing and established methods for improving study recruitment and retention in multi-visit studies.

Conflicts of Interest

None declared.

References

1. Cooley ME, Sarna L, Brown JK, Williams RD, Chernecky C, Padilla G, et al. Challenges of recruitment and retention in multisite clinical research. Cancer Nurs 2003 Oct;26(5):376-384. [Medline: 14710799]
2. Epstein WM. Randomized controlled trials in the human services. Soc Work Res Abstr 1993 Sep 01;29(3):3-10. [doi: 10.1093/swra/29.3.3]
3. Good M, Schuler L. Subject retention in a controlled clinical trial. J Adv Nurs 1997 Aug;26(2):351-355. [Medline: 9292370]
4. Siddiqui O, Flay BR, Hu FB. Factors affecting attrition in a longitudinal smoking prevention study. Prev Med 1996;25(5):554-560. [doi: 10.1006/pmed.1996.0089] [Medline: 8888323]
5. Gul R, 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]
6. Patel M, Doku V, Tennakoon L. Challenges in recruitment of research participants. Adv Psychiatr Treat 2018 Jan 02;9(3):229-238. [doi: 10.1192/apt.9.3.229]
7. Chang R, Falconer J, Stulberg SD, Arnold WJ, Dyer AR. Prerandomization: an alternative to classic randomization. The effects on recruitment in a controlled trial of arthroscopy for osteoarthrosis of the knee. J Bone Joint Surg Am 1990 Dec;72(10):1451-1455. [Medline: 2254351]
8. Gross D, Julion W, Fogg L. What motivates participation and dropout among low-income urban families of color in a prevention intervention? Fam Relat 2001 Jul;50(3):246-254. [doi: 10.1111/j.1741-3729.2001.00246.x]
9. Mason MJ. Methods, plainly speaking: A review of procedural and statistical methods for handling attrition and missing data in clinical research. Meas Eval Couns Dev 1999;32(2):111.
10. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. BMJ open 2013;3(e002360):1-24 [FREE Full text] [doi: 10.1002/14651858.MR000013.pub5]
11. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health 2011;11(249):1-12. [doi: 10.1186/1471-2458-11-249] [Medline: 21504610]
12. Bentley JP, Thacker PG. The influence of risk and monetary payment on the research participation decision making process. J Med Ethics 2004 Jun;30(3):293-298 [FREE Full text] [Medline: 15173366]
13. Halpern S, Karlawish JH, Casarett D, Berlin JA, Asch DA. Empirical assessment of whether moderate payments are undue or unjust inducements for participation in clinical trials. Arch Intern Med 2004 Apr 12;164(7):801-803. [doi: 10.1001/archinte.164.7.801] [Medline: 15078651]
14. Scott C. A replicable model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. Drug Alcohol Depend 2004 Apr 09;74(1):21-36 [FREE Full text] [doi: 10.1016/j.drugalcdep.2003.11.007] [Medline: 15072804]
15. Kahler C, Metrik J, Spillane N, Tidey J, Leventhal A, Mckee S, et al. Alcohol dependence moderates the effects of alcohol on smoking lapse in a laboratory analogue model: s264. Alcoholism: Clinical & Experimental Research 2012;36:79A.
16. Kahler C, Metrik J, Spillane NS, Day A, Leventhal AM, McKee SA, et al. Acute effects of low and high dose alcohol on smoking lapse behavior in a laboratory analogue task. Psychopharmacology (Berl) 2014 Dec;231(24):4649-4657 [FREE Full text] [doi: 10.1007/s00213-014-3613-3] [Medline: 24858377]
17. Smith L, McNamara PJ, King AC. Optimizing follow-up and study retention in the 21st century: advances from the front line in alcohol and tobacco research. Drug Alcohol Depend 2017 Dec 01;175:171-178 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.01.045] [Medline: 28437721]
18. Robinson K, Dinglas VD, Sukrithan V, Yalamanchilli R, Mendez-Tellez PA, Dennison-Himmelfarb C, et al. Updated systematic review identifies substantial number of retention strategies: using more strategies retains more study participants. J Clin Epidemiol 2015 Dec;68(12):1481-1487 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.013] [Medline: 26186981]
19. Leavens EL, Morgan TL, Brett EI, Lee K, Son J, Molina N, et al. Concurrent alcohol use and waterpipe tobacco smoking: smoking topography, toxicant exposure, and abuse liability. Nicotine Tob Res 2019:- (forthcoming).
20. Griffiths R, Troisi JR, Silverman K, Mumford GK. Multiple-choice procedure: an efficient approach for investigating drug reinforcement in humans. Behav Pharmacol 1993 Feb;4(1):3-13. [Medline: 11224166]

Abbreviations

MCP: multiple-choice procedure
SMS: short message service

Edited by G Eysenbach; submitted 30.05.18; peer-reviewed by E Warner, A Cyr; comments to author 17.08.18; revised version received 29.10.18; accepted 10.12.18; published 08.04.19.

Please cite as:
Leavens ELS, Stevens EM, Brett EI, Molina N, Leffingwell TR, Wagener TL
Use of Rideshare Services to Increase Participant Recruitment and Retention in Research: Participant Perspectives
J Med Internet Res 2019;21(4):e11166
URL: https://www.jmir.org/2019/4/e11166/
doi:10.2196/11166
PMID:30958268

Original Paper

Effects of Mock Facebook Workday Comments on Public Perception of Professional Credibility: A Field Study in Canada

Cynthia Weijs[1,2], MPH, PhD; Jason Coe[2], DVM, PhD; Serge Desmarais[3], PhD; Shannon Majowicz[4], PhD; Andria Jones-Bitton[2], DVM, PhD

[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada

[3]Department of Psychology, University of Guelph, Guelph, ON, Canada

[4]Faculty of Applied Health Sciences, School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Cynthia Weijs, MPH, PhD
Department of Community Health Sciences
Cumming School of Medicine
University of Calgary
TRW3, 3280 Hospital Dr. NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 4299
Email: cynthia.weijs@ucalgary.ca

Abstract

Background: There is considerable discussion of risks to health professionals’ reputations and employment from personal social media use, though its impacts on professional credibility and the health professional-client relationship are unknown.

Objective: The aim of this study was to test the extent to which workday comments posted to health professionals’ personal Facebook profiles influence their credibility and affect the professional-client relationship.

Methods: In a controlled field study, participants (members of the public) reviewed randomly assigned mock Facebook profiles of health professionals. The 2×2×2 factorial design of mock profiles included gender (female/male), health profession (physician/veterinarian), and workday comment type (evident frustration/ambiguous). Participants then rated the profile owner’s credibility on a visual analog scale. An analysis of variance test compared ratings. Mediation analyses tested the importance of credibility ratings on participants’ willingness to become a client of the mock health professional.

Results: Participants (N=357) rated health professionals whose personal Facebook profile showed a comment with evident frustration rather than an ambiguous workday comment as less credible (*P*<.001; mean difference 11.18 [SE 1.28]; 95% CI 8.66 to 13.70). Furthermore, participants indicated they were less likely to become clients of the former when they considered credibility (standardized beta=.69; *P*<.001). Credibility explained 86% of the variation in the relationship between the type of workday comment and the participant’s willingness to become a client of the health professional.

Conclusions: This study provides the first evidence of the impact of health professionals’ personal online disclosures on credibility and the health relationship. Public perceptions about professionalism and credibility are integral to developing the evidence base for e-professionalism guidelines and encouraging best practices in social media use.

(J Med Internet Res 2019;21(4):e12024) doi:10.2196/12024

KEYWORDS

social media; professionalism; trust; professional-patient relations; medical education

Introduction

Changing Context of Professionalism

Professionalism is important to the health professional-client relationship (hereafter, the health relationship) [1]. Perceptions of professional credibility, generally defined as caring, competence, and trustworthiness [2], underpin clients' sense of professionalism [1,3]. Typically, such impressions are formed during health interactions, though widespread internet use has introduced new contexts for impression formation. In one survey, 80% of US internet users sought health information online, with 44% specifically searching for information about health professionals [4]. Social media have garnered attention as being a risk to health professionals who use such sites in their personal lives. One site, Facebook, may reveal significant personal information that health professionals would be unlikely to share in the course of a client interaction. Given Facebook's immense popularity with the public [5,6] and health professionals [7-11], and the call for evidence about public impressions of credibility developed in this new context [12], we explored public impressions of health professionals who use Facebook in their private lives.

Reputation Risk When Private and Professional Lives Mix Online

An important issue is the need to strike a balance between health professionals' right to use Facebook as private citizens and their duty to protect professional reputation and health relationships [13,14]. Reputation damage, as evidenced by employment loss and regulatory discipline [15,16], has occurred. However, the consequence to the health relationship from health professionals' personal use of Facebook, or whether patients notice these behaviors or see them as inappropriate, is complex and less evidenced. Yet, 49 out of 72 (68%) of medical students surveyed in a social media and professionalism intervention reported the mixing of their personal and professional lives on social media sites as a key concern [17].

Normative beliefs about online or e-professionalism among some practicing health professionals and trainees differ considerably [7-9,11,18], perhaps mirroring the professionalism transition seen in the academic and medical education literature, from the old (unquestioned authority, paternalistic, and emotionally distant) to the new (client-centered, collaborative, and emotionally present) [19]. Some see e-professionalism as *a matter of taste*, overly rigid, individual, and impossible to attain [10]. Although it has been a decade since the advent of social media, very little evidence exists to inform social media guidelines about the interplay between personal online disclosures and professional trust and credibility [12] or whether expectations of e-professionalism might differ between health professional groups.

Objectives

We explored the impact of workday comments that blur the line between health professionals' private and professional lives as they are common and thought to have an impact on perceived professionalism [18,20-22]. We aimed to test (1) whether members of the public independently notice workday comments on Facebook profiles belonging to physicians or veterinarians, (2) whether such comments influence perceptions of professionalism (operationalized as credibility evaluations), and (3) the impact on the willingness of participants to engage with those health professionals (ie, become a client).

Methods

Informed Consent and Recruitment

The study was approved by the Research Ethics Board at the University of Guelph. Informed consent consisted of a welcome message and a letter of information on the landing page of a Web-based questionnaire. The sample was a convenience sample. Participants were recruited in person in communities across Southwestern Ontario or online, through local social media (ie, advertisements about the study on Facebook and Kijiji, a popular online classifieds site in Canada). Members of the public were approached in public spaces (eg, farmers' markets, public trailheads, parks, soccer fields, and sports games) to take part in a study described as aiming to understand people's first impressions of personality from online content. Participants who showed interest in participating were given a card with the researchers' contact information and URL for the survey. The survey was open and voluntary. To control potential social desirability bias, no mention was made of health professionals [23]. Participants were encouraged to share the online survey link through email and Facebook. Data were captured automatically by a Web survey application, Fluidsurvey, and downloaded onto password protected and encrypted computers. After completing the survey, participants submitted their email contact information on a separate webpage to have a 1 in 350 chance of winning an Apple iPad.

Survey Design

The first of 3 survey sections presented a randomly assigned mock Facebook profile in PDF format for participants' review (a mock profile is available to readers from the authors upon request). The second survey section presented an adapted credibility scale on which participants rated the mock profile owner [2]. Finally, 16 demographic questions and 1 question asking participants about their willingness to become a client of the profile owner (a physician or veterinarian) were presented.

Table 1. Adapted credibility scale showing paired anchor characteristics for competence, caring, and trustworthiness (Cronbach alpha for the scale=.88). Two changes were made to the original validated scale[a,b].

Competence	Caring	Trustworthiness
Novice/Expert	Insensitive/Sensitive	Phoney/Genuine
Unintelligent/Intelligent	Not understanding/Understanding	Unethical/Ethical
Incompetent/Competent	Self-centered/Unselfish	Untrustworthy/Trustworthy
Uninformed/Informed	Unconcerned/Concerned	Dishonest/Honest
Dim[a]/Bright	Uncaring/Caring	Immoral/Moral
Untrained/Trained	Doesn’t care about me/Cares about me[b]	Dishonourable/Honourable

[a] *Dim* replaced the original adjective, *Stupid*.

[b] This pair was removed from the scale owing to irrelevance to our study scenarios.

Profiles were made to simulate a real Facebook profile, although without working links and with names and headshot photos contributed by members of the research team with permission. Profiles were created using a 2×2×2 full factorial design, with the following attributes: mock profile owner gender (female/male), health discipline (physician/veterinarian), and workday comment type (evident frustration/ambiguous). Profile comments (ie, status updates) were modeled after actual status updates from previous studies of health professionals’ personal use of Facebook to authentically portray workday comments as they may be encountered online [18,20-22]. Profile photos of Sarah or Chad, the mock profile owners, were of a sunset, to control for potential bias related to facial features upon first impression. Several workday comments were pilot tested within the research team and with members of the public for content validity. In our post pilot test debrief, we asked participants what they understood the paired adjectives to mean. They all identified at least one of the exact labels of the scale (competence, caring, trustworthiness, or credibility) or a synonym of the scale (eg, professionalism). Additionally, we inquired about their views on the piloted comments as we aimed for a moderate comment in terms of emotional valence. On the basis of their feedback, we chose the following evident workday frustration comment:

> *What is it with some people?? I know I only went through 9 years of university...but really, I know what I’m talking about...yeesh!!!* [Sarah or Chad]

We also chose the following ambiguous workday comment:

> *Started with new electronic patient charts today...interesting experience for sure.* [Sarah or Chad]

Each participant reviewed the content of 1 randomly assigned mock Facebook profile without time constraints. On the next page, they completed the 17 personality ratings, an adapted credibility scale which included 3 subscales: competence (6 items), caring (5 items), and trustworthiness (6 items; Table 1) [2]. The scale originates from a widely cited study by McCroskey and Teven that validated a credibility scale composed of 3 related dimensions—goodwill (caring), competence, and trustworthiness [2]. We modified 2 items in the scale, replacing the original adjective, stupid, with dim in the competence subscale and removing 1 set of adjectives that was not relevant to our scenarios in the goodwill or caring subscale (Table 1).

Each item was represented on a visual analog scale, anchored at each end with an adjective describing less (eg, novice) or more (eg, expert) of the trait being assessed (eg, competence). Consistent with first impression research, participants were asked to move fairly quickly through all randomly presented rating scales, sliding the cursor to the place on the line between the 2 adjectives that they felt best represented the profile owner [24,25]. The question assessing a participant’s willingness to become a client of the mock profile owner was similarly represented, anchored by very unlikely and very likely.

Numeric gradations for all visual analog scales (0 to 100) were hidden to access participants’ immediate impressions. Only one response was possible, and all items had a nonresponse option of nonmovement from the middle position on the scale (neutral). Respondents were able to review and change their answers before moving to the next page and could not move backward in the survey so as to ensure collection of their *first* impressions. Duplicate entries were identified through a combination of internet protocol address and demographic details. All submitted surveys were included in the analysis, except those that were less than half complete or had fewer than 2 credibility subscales completed. The survey was pretested on a group of 10 individuals known to the researchers who were debriefed upon completing the survey about usability, clarity, functionality, and content validity of the credibility scale.

Statistical Analysis

Univariable linear regression identified unconditional associations between the outcome variable (credibility rating) and covariate or control variables including the following: participant is a Facebook profile owner (no; yes); recruitment (in person; online); participant age (years); participant gender (female; male); annual household income before taxes (low ≤Can $ 74,999; high ≥Can $ 75,000); and participant’s profile viewing and rating times (in seconds). Covariates having a *P* value <.20 with the outcome variable were included in an omnibus analysis of variance (ANOVA), along with the 3 manipulated mock profile factors (profile owner gender, health discipline, and workday comment type), to determine credibility differences based on comment type.

Sequential linear regression determined the extent to which impressions gleaned from the online content of a personal Facebook profile impact the health relationship [26]. Specifically, credibility was assessed for its role as a mediator (intervening variable) in the relationship between profile workday comment type and the participant's willingness to become a client of the profile owner. As above, univariable linear regression was employed to reduce variables eligible for inclusion in the model. In Step 1, the 3 manipulated variables in our study and any eligible covariates were entered into the model. In Step 2, credibility was added to the model. Mediation was supported if the intervening variable maintained a significant standardized beta coefficient when all other covariates were controlled [26]. Full mediation as per Tabachnick and Fidell was considered to have occurred when beta coefficients for significant variables in Step 1 were no longer significant in Step 2 [26]. Effect size, as per Cohen, and 95% CIs for the sequential regression model were calculated [27,28]. All statistical analyses were carried out in SPSS v23 (SPSS, Inc) using a significance level of *P*<.05.

Results

User Statistics

The total number of people invited to participate in the survey is unknown. A total of 6 surveys were removed owing to insufficient responses, leaving 357 usable surveys. Not all participants answered all questions, thus samples vary as noted below. A missing value analysis revealed that values were missing completely at random [26]. Participants were, on average, 40 years old (median 40 (SD 12.4); range 18-73) and 226 out of 355 (63.7%) were female. Participants' income and education are shown in Table 2. Most participants (289/348, 83.0%) reported having a personal Facebook profile, and among those, 159 out of 285 (55.8%) reported checking it daily or several days per week (57/285, 20.0%). On average, participants viewed the mock Facebook profiles for 87 seconds (median 74 (SD 62); range 6.23-592 seconds), took 124.00 seconds (median 110 (SD 61.5); range 32-652 seconds) to rate the mock profile owners, and took 7.15 min (median 6.60 (SD 2.54); range 2.18-17.53 min) to complete the entire survey.

Differences in Credibility Ratings of Health Professionals Having Either Evident Frustration or Ambiguous Comments on Mock Facebook Profiles

The following variables were included in the omnibus ANOVA: workday comment type, mock profile owner discipline and gender, participant age and gender, participant profile viewing time, and recruitment venue (in person, online). The ANOVA results yielded only one statistically significant main effect—workday comment type, $F_{1,339}$=64.03; *P*<.001—that is, mock Facebook profile owners with an evident workday frustration comment were rated as significantly less credible than were those with an ambiguous workday comment (mean difference 11.2 (SE 1.3); 95% CI 8.66 to 13.70). The average credibility rating for evident profiles was 56.7 out of 100, whereas for ambiguous profiles it was 67.9 out of 100. The model-adjusted R^2 was 0.188, indicating 19% of the variation in credibility ratings was due to the workday comment type.

Understanding the Real-World Effect of Workday Comments and Credibility Ratings on Health Relationships

Sequential regression supported mediation, meaning that the effect of workday comment type (evident frustration versus ambiguous) on participants' willingness to become a client of health professional profile owners was mediated by credibility ratings (Tables 3 and 4). Cohen f^2 effect size for the model was large (0.86; 95% CI for f^2 0.62 to 1.18), meaning 86% of the variation in participants' willingness to become a client of mock health professional profile owners was due to the addition of credibility rating to the model. Participants perceived the profile comment as a reflection of credibility, which subsequently impacted their willingness to become a client of the profile owner (higher credibility ratings associated with more willingness).

Table 2. Number (n) and proportion (%) of study participants within pretax annual household income and completed education categories.

Demographic characteristic	n (%)
Pretax annual household income in Can $	**n=335**
<$30,000	50 (15)
$30,000-$49,999	42 (13)
$50,000-$74,999	75 (22)
$75,000-$99,999	71 (21)
≥$100,000	97 (29)
Level of completed education	**n=357**
Completed or have some high school	52 (14)
College diploma	66 (18)
University degree	118 (33)
Professional degree	48 (13)
Graduate degree	72 (22)

Table 3. Step 1 of the sequential regression examining credibility as a mediator in the relationship between workday comment type and willingness to become a patient of the health professional for 318 study participants (R^2=0.19; adjusted R^2=0.17; and $F_{7, 310}$=10.44).

Variable	Statistics	
	Standardized beta coefficient	*P* value
Participant age	−.03	.51
Participant income (ref[a] ≤Can $ 74,999)	−.08	.17
Mock profile owner gender (ref female)	−.10	.05
Health discipline (ref physician)	.65	.21
Ambiguous workday comment (ref evident frustration)	.39	<.001
Online recruitment (ref in person)	−.02	.68
Scenario view time (in seconds)	−.08	.12
Credibility rating	—[b]	—

[a] Ref: referent.

[b] Variable not included in the model in Step 1.

Table 4. Step 2 of the sequential regression examining credibility as a mediator in the relationship between workday comment type and willingness to become a patient of a physician or client of a veterinarian for 318 study participants (R^2=0.57; adjusted R^2=0.56; change in R^2=0.38; and $F_{1,309}$=267.48).

Variable	Statistics	
	Standardized beta coefficient	*P* value
Participant age	−.02	.70
Participant income (ref[a] ≤Can $74,999)	−.05	.18
Mock profile owner gender (ref female)	−.09	.02
Health discipline (ref physician)	.06	.11
Ambiguous workday comment (ref evident frustration)	.10	.03
Online recruitment (ref in person)	.06	.10
Scenario view time (in seconds)	−.06	.12
Credibility rating	.69	<.001

[a] Ref: referent.

Discussion

Principal Findings

Facebook functions as a space to share information and access social support [29,30], including sharing workday frustrations [31]. Health professionals, regulators, and educators aim to better understand health professionals' personal use of social media and networking sites, especially Facebook, given its ubiquity in the personal lives of the general public and of health professionals [5-11]. This field study provides some of the first evidence that the blurring between private and professional lives that is inherent on health professionals' personal Facebook profiles warrants the attention of health professional communities. Participants independently noticed a subtle workday frustration comment on a mock health professional's personal Facebook account and perceived that individual as less credible than a mock health professional with an ambiguous workday comment. Participants also indicated less willingness to engage professionally with the profile owners who posted evident versus ambiguous frustration comments.

We found a small, though not statistically significant, difference between participant credibility ratings of physicians and veterinarians. Given the different health contexts for physicians and veterinarians, the findings of this study suggest that the public has expectations of online professionalism that warrant further exploration across a range of health professions to broaden our understanding of credibility evaluations in relation to and beyond those studied here.

Comparison With Previous Work

Our results seem consistent with a conventional view of professionalism and suggest that current cautions around health professionals' personal use of Facebook are not unreasonable [32]. Yet, mock profile owners in our study were rated somewhat less harshly (ie, still rated above the midpoint of the professionalism scale at 57 out of 100) than those in a similar study [33]; however, study design features may account for this difference. Jain et al had employees of a US university rate appropriateness (their measure of professionalism) of workday comments about patients made by medical students on a Likert scale (from 1-very inappropriate to 5-very appropriate); the average score was 1.88 [33]. Participants in that study were informed that the content was publicly available and searchable on Facebook and were asked to rate behaviors that were circled in red. In addition, workplace comments were combined for analysis, blending milder and more severe comments, all of which may have contributed to harsher ratings [33]. In contrast, we aimed to control socially desirable responses by having participants review the profile for as long as they wished, leaving them to independently notice both the workday comment and that the individual was a health professional.

In another study, Clyde et al suggest that clients may be motivated to search for health professionals online to supplement information available on a workplace website [34]. They found that Facebook profiles belonging to mock physicians with healthy personal information (eg, reading and hiking) were perceived as significantly more professional than profiles that limited content solely to information about mock physicians' education and current practice. Clyde et al interpreted this to mean that access to personal information allowed participants to better judge physicians' professionalism attributes [34]. Their finding was unexpected [34], perhaps because conventionally, personal self-disclosure (even of a positive nature) by physicians is thought to lower professional boundaries and is largely perceived as having a detrimental effect on the health relationship [35,36]. However, as with this study, their findings highlight the importance that health professionals' personal disclosures online may have with patients.

Strengths and Limitations

Key strengths of this field study were the random presentation of mock profiles and the realistic context of the survey design, which parallels the way a health professional's Facebook information may be discovered online. The use of indirect questions to assess judgments of health professionals [37] and anonymity in survey completion [38] likely limited social desirability responses. The proportion of females (62%) to males in this study was more representative than in traditional survey methods, where women often comprise 71% to 77% of the participants [39], and the average age of participants in this study (40 years) is similar to the average age of a social media user (38 years) [29].

Nevertheless, choosing a health care provider involves more factors than were measured in this study, such as location and word-of-mouth referrals. Facial gestures and physical traits also influence impressions [40], as do characteristics of the trustor [41], comments posted by friends, and personality traits of the trustee [42,43], which were not examined here.

Implications for Practice and Social Media Guidelines

In a recent exploration of dentists' and patients' attitudes toward social media use, 36% of patients reported searching for their dentist or doctor on social media. Other findings highlighted patients' desire for more information about their health professionals (eg, by reading online reviews and qualifications) [44]. These findings underscore the near impossibility for health professionals in defending strict boundaries between their private and professional selves online but, rather, suggest the value of further research to better understand the impacts of unintentional *patient communication* of this type.

We conclude that health professionals, especially early-career professionals who have yet to build their professional reputation, should be mindful of their credibility with the public when using Facebook in their private lives. Adhering to social media guidelines that take a traditional or conventional view of professionalism is a reasonable first step, especially around workday comments. Although confidentiality breeches are more serious and have been a focus of attention for regulators and educators, they are rare [12,18], whereas, workday comments are common on Facebook [12,18-20,22]. We acknowledge that such a recommendation is associated with significant burden for health professionals, who have a right to access social support on Facebook when acting as private citizens. However, these rights need to be balanced with their professional obligation to maintain positive health relationships [13].

Conclusions

Although a fuller understanding of e-professionalism is ongoing in this field, given the profound effect of digital technologies on society, social media guidelines should incorporate early evidence to prevent credibility damage because (1) trust, once broken, is difficult to rebuild [45]; (2) Facebook is seemingly rooted in society, with 1 in 5 Americans using social media for health care decisions (94% of whom reported Facebook as their trusted source) [46]; and (3) recent evidence suggests that health professionals do indeed *friend* patients [44]. This study underscores that the blurring of professional and personal lives on social media is important to understand and manage. Better elucidating patients' differing perspectives of e-professionalism around various workday comments and identifying contextual influences, such as views about whether Facebook is private or public, will better define expectations of e-professionalism, which may in turn alleviate some of the concerns and reticence among health professionals to participate in social media.

Furthermore, an untapped area of health communication research includes the evaluation of health professionals use of social media skills in the public domain, that is, when, as private citizens, they participate in broader public online discussion about health matters within their professional domain (eg, posting comments in the comments section in news articles about vaccination).

Acknowledgments

CW's stipend was funded by the Canadian Institutes of Health Research under grant number 599547.

Conflicts of Interest

None declared.

References

1. Cruess SR, Johnston S, Cruess RL. "Profession": a working definition for medical educators. Teach Learn Med 2004;16(1):74-76. [doi: 10.1207/s15328015tlm1601_15] [Medline: 14987179]
2. McCroskey JC, Teven JJ. Goodwill: a reexamination of the construct and its measurement. Commun Monogr 2009 Jun 02;66(1):90-103. [doi: 10.1080/03637759909376464]
3. Shay LA, Dumenci L, Siminoff LA, Flocke SA, Lafata JE. Factors associated with patient reports of positive physician relational communication. Patient Educ Couns 2012 Oct;89(1):96-101 [FREE Full text] [doi: 10.1016/j.pec.2012.04.003] [Medline: 22554386]
4. Fox S. Pew Research Center. 2011 Feb 01. Health topics URL: http://pewinternet.org/Reports/2011/HealthTopics.aspx [accessed 2018-08-23] [WebCite Cache ID 71tHGIIbn]
5. Christofides E, Muise A, Desmarais S. Hey Mom, What's on Your Facebook? Comparing Facebook disclosure and privacy in adolescents and adults. Soc Psychol Pers Sci 2011 May 17;3(1):48-54. [doi: 10.1177/1948550611408619]
6. Christofides E, Muise A, Desmarais S. Information disclosure and control on Facebook: are they two sides of the same coin or two different processes? Cyberpsychol Behav 2009 Jun;12(3):341-345. [doi: 10.1089/cpb.2008.0226] [Medline: 19250020]
7. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
8. 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]
9. Coe JB, Weijs CA, Muise A, Christofides E, Desmarais S. Understanding veterinary students' use of and attitudes toward the social networking site, Facebook, to assist in developing curricula to address online professionalism. J Vet Med Educ 2012;39(3):297-303. [doi: 10.3138/jvme.0212-016R] [Medline: 22951465]
10. Ross S, Lai K, Walton JM, Kirwan P, White JS. "I have the right to a private life": medical students' views about professionalism in a digital world. Med Teach 2013 Oct;35(10):826-831. [doi: 10.3109/0142159X.2013.802301] [Medline: 23826730]
11. Weijs CA, Coe JB, Muise A, Christofides E, Desmarais S. Reputation management on facebook: awareness is key to protecting yourself, your practice, and the veterinary profession. J Am Anim Hosp Assoc 2014;50(4):227-236. [doi: 10.5326/JAAHA-MS-6069] [Medline: 24855091]
12. 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]
13. 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]
14. Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age. Current Pharm Teach Learn 2009 Dec;1(2):66-70. [doi: 10.1016/j.cptl.2009.10.001]
15. Conaboy C. The Boston Globe. 2011 Apr 20. For doctors, social media a tricky case URL: http://www.webcitation.org/71sR7xWAq [accessed 2018-08-22] [WebCite Cache ID 71sR7xWAq]
16. Griffith R, Tengnah C. District nurses' use of social networking sites: caution required. Br J Community Nurs 2011 Sep;16(9):455-457. [doi: 10.12968/bjcn.2011.16.9.455] [Medline: 22067956]
17. 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]
18. Black EW, Thompson LA, Duff WP, Dawson K, Saliba H, Black NM. Revisiting social network utilization by physicians-in-training. J Grad Med Educ 2010 Jun;2(2):289-293 [FREE Full text] [doi: 10.4300/JGME-D-10-00011.1] [Medline: 21975635]
19. Borgstrom E, Cohn S, Barclay S. Medical professionalism: conflicting values for tomorrow's doctors. J Gen Intern Med 2010 Dec;25(12):1330-1336 [FREE Full text] [doi: 10.1007/s11606-010-1485-8] [Medline: 20740324]
20. Coe JB, Weijs CA, Muise A, Christofides E, Desmarais S. Teaching veterinary professionalism in the Face(book) of change. J Vet Med Educ 2011;38(4):353-359. [doi: 10.3138/jvme.38.4.353] [Medline: 22130411]
21. Weijs CA, Coe JB, Christofides E, Muise A, Desmarais S. Facebook use among early-career veterinarians in Ontario, Canada (March to May 2010). J Am Vet Med Assoc 2013 Apr 15;242(8):1083-1090. [doi: 10.2460/javma.242.8.1083] [Medline: 23547671]
22. Thompson LA, Dawson K, Ferdig R, Black EW, Boyer J, Coutts J, et al. The intersection of online social networking with medical professionalism. J Gen Intern Med 2008 Jul;23(7):954-957 [FREE Full text] [doi: 10.1007/s11606-008-0538-8] [Medline: 18612723]

23. Edwards AL. The Social Desirability Variable in Personality Assessment and Research. Worcester, Massachusetts, United States: American Psychological Association; 1958:610-611.
24. Heise D. The Semantic differential attitude research. In: Summers GF, editor. Attitude Measurement. Oregon: Kershaw Publishing Co Ltd; 1970:235-253.
25. Gerich J. Visual analogue scales for mode-independent measurement in self-administered questionnaires. Behav Res Methods 2007 Nov;39(4):985-992. [Medline: 18183916]
26. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th edition. Boston, MA: Pearson; 2013:-.
27. Soper D. Daniel Soper. 2015. Effect Size Calculator for Hierarchical Multiple Regression URL: https://www.danielsoper.com/statcalc/default.aspx [accessed 2018-08-23] [WebCite Cache ID 71tEyi9Uk]
28. Soper D. Daniel Soper. 2015. f-square effect size confidence interval calculator URL: https://www.danielsoper.com/statcalc/calculator.aspx?id=75 [accessed 2018-08-23] [WebCite Cache ID 71tFCOR9H]
29. Hampton K, Goulet L, Rainie L, Purcell K. Pew Research Center. 2011 Jun 16. Social networking sites and our lives URL: http://www.pewinternet.org/2011/06/16/social-networking-sites-and-our-lives/ [accessed 2018-08-23] [WebCite Cache ID 71tGULztR]
30. Ellison NB, Vitak J, Gray R, Lampe C. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. J Comput Mediat Commun 2014 Jul 01;19(4):855-870. [doi: 10.1111/jcc4.12078]
31. Farnan JM, Paro JA, Higa J, Edelson J, Arora VM. The YouTube generation: implications for medical professionalism. Perspect Biol Med 2008;51(4):517-524. [doi: 10.1353/pbm.0.0048] [Medline: 18997354]
32. Mayer MA, Leis A, Mayer A, Rodriguez-Gonzalez A. How medical doctors and students should use Social Media: a review of the main guidelines for proposing practical recommendations. Stud Health Technol Inform 2012;180:853-857. [Medline: 22874313]
33. 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 [FREE Full text] [doi: 10.1111/medu.12282] [Medline: 24528398]
34. Clyde JW, Domenech Rodríguez MM, Geiser C. Medical professionalism: an experimental look at physicians' Facebook profiles. Med Educ Online 2014;19:23149 [FREE Full text] [doi: 10.3402/meo.v19.23149] [Medline: 24947922]
35. McDaniel SH, Beckman HB, Morse DS, Silberman J, Seaburn DB, Epstein RM. Physician self-disclosure in primary care visits: enough about you, what about me? Arch Intern Med 2007 Jun 25;167(12):1321-1326. [doi: 10.1001/archinte.167.12.1321] [Medline: 17592107]
36. Beach MC, Roter D, Rubin H, Frankel R, Levinson W, Ford DE. Is physician self-disclosure related to patient evaluation of office visits? J Gen Intern Med 2004 Sep;19(9):905-910 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.40040.x] [Medline: 15333053]
37. Lount RB, Pettit NC. The social context of trust: the role of status. Organ Behav Hum Decis Process 2012 Jan;117(1):15-23. [doi: 10.1016/j.obhdp.2011.07.005]
38. Denissen JJ, Neumann L, van Zalk M. How the internet is changing the implementation of traditional research methods, people’s daily lives, and the way in which developmental scientists conduct research. Int J Behav Dev 2010 Oct 27;34(6):564-575. [doi: 10.1177/0165025410383746]
39. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004 Mar;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
40. Krumhuber E, Manstead AS, Cosker D, Marshall D, Rosin PL, Kappas A. Facial dynamics as indicators of trustworthiness and cooperative behavior. Emotion 2007 Nov;7(4):730-735. [doi: 10.1037/1528-3542.7.4.730] [Medline: 18039040]
41. Viggiano MP, Marzi T. Chapter 8: Context and social effects on face recognition. In: Balcetis E, Lassiter DG, editors. The Social Psychology of Visual Perception. New York, NY: Psychology Press; 2010:171-200.
42. Tong ST, van der Heide B, Langwell L, Walther JB. Too much of a good thing? The relationship between number of friends and interpersonal impressions on facebook. J Comput Mediat Commun 2008;13(3):531-549. [doi: 10.1111/j.1083-6101.2008.00409.x]
43. Utz S, Tanis M, Vermeulen I. It is all about being popular: the effects of need for popularity on social network site use. Cyberpsychol Behav Soc Netw 2012 Jan;15(1):37-42. [doi: 10.1089/cyber.2010.0651] [Medline: 21988765]
44. Parmar N, Dong L, Eisingerich AB. Connecting with your dentist on Facebook: patients' and dentists' attitudes towards social media usage in dentistry. J Med Internet Res 2018 Jun 29;20(6):e10109 [FREE Full text] [doi: 10.2196/10109] [Medline: 29959108]
45. Rempel JK, Holmes JG, Zanna MP. Trust in close relationships. J Personal Soc Psychol 1985;49(1):95-112. [doi: 10.1037/0022-3514.49.1.95]
46. National Research Corporation. 2011. 1 in 5 Americans Use Social Media for Health Care Information URL: http://hcmg.nationalresearch.com/public/News.aspx?ID=9 [accessed 2018-08-21] [WebCite Cache ID 71tH2VgLU]

Abbreviations

ANOVA: analysis of variance

Edited by G Eysenbach; submitted 23.08.18; peer-reviewed by J Cain, B Bie, C Lammert, S Manca, A Kulanthaivel; comments to author 23.01.19; accepted 17.02.19; published 18.04.19.

Please cite as:
Weijs C, Coe J, Desmarais S, Majowicz S, Jones-Bitton A
Effects of Mock Facebook Workday Comments on Public Perception of Professional Credibility: A Field Study in Canada
J Med Internet Res 2019;21(4):e12024
URL: http://www.jmir.org/2019/4/e12024/
doi:10.2196/12024
PMID:30998223

Original Paper

Privacy-Preserving Record Grouping and Consent Management Based on a Public-Private Key Signature Scheme: Theoretical Analysis and Feasibility Study

Stephan Jonas[1], Dr rer medic; Simon Siewert[2]; Cord Spreckelsen[2], PD, Dr rer nat

[1]Department of Informatics, Technical University of Munich, Garching, Germany

[2]Department of Medical Informatics, Uniklinik Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany

Corresponding Author:
Stephan Jonas, Dr rer medic
Department of Informatics
Technical University of Munich
Boltzmannstrasse 3
Garching, 85748
Germany
Phone: 49 49 089 289 18
Email: jonas@in.tum.de

Abstract

Background: Clinical and social trials create evidence that enables medical progress. However, the gathering of personal and patient data requires high security and privacy standards. Direct linking of personal information and medical data is commonly hidden through pseudonymization. While this makes unauthorized access to personal medical data more difficult, a centralized pseudonymization list can still pose a security risk. In addition, medical data linked via pseudonyms can still be used for data-driven reidentification.

Objective: Our objective was to propose a novel approach to pseudonymization based on public-private key cryptography that allows (1) decentralized patient-driven creation and maintenance of pseudonyms, (2) 1-time pseudonymization of each data record, and (3) grouping of patient data records even without knowing the pseudonymization key.

Methods: Based on public-private key cryptography, we set up a signing mechanism for patient data records and detailed the workflows for (1) user registration, (2) user log-in, (3) record storing, and (4) record grouping. We evaluated the proposed mechanism for performance, examined the potential risks based on cryptographic collision, and carried out a threat analysis.

Results: The performance analysis showed that all workflows could be performed with an average runtime of 0.057 to 42.320 ms (user registration), 0.083 to 0.606 ms (record creation), and 0.005 to 0.198 ms (record grouping) depending on the chosen cryptographic tools. We expected no realistic risk of cryptographic collision in the proposed system, and the threat analysis revealed that 3 distinct server systems of the proposed setup had to be compromised to allow access to combined medical data and private data. However, this would still allow only for data-driven deidentification. For a full reidentification, all 3 trial servers and all study participants would have to be compromised. In addition, the approach supports consent management, automatically anonymizes the data after trial closure, and provides basic mechanisms against data forging.

Conclusions: The proposed approach has a high security and privacy level in comparison with traditional centralized pseudonymization approaches and does not require a trusted third party. The only drawback in comparison with central pseudonymization is the directed feedback of accidental findings to individual participants, as this is not possible with a quasi-anonymous storage of patient data.

(J Med Internet Res 2019;21(4):e12300) doi:10.2196/12300

KEYWORDS

asymmetric cryptography; public-private key; long-term trials; clinical trials as topic; data anonymization; pseudonymization

Introduction

Background

Medical progress relies on evidence from trials involving patients, healthy participants, or both. Many relevant study designs (especially cohort studies) require long-term efforts, which need to merge data for each participant. The same holds true for studies that capture and then integrate data from heterogeneous data sources or from different locations to describe individual participants. While handling sensitive health data, such studies require the highest standards of data privacy and data security. Specifically, they have to ensure that sensitive data cannot be traced back to individual participants by unauthorized access. Obviously, these two requirements—enabling record grouping for individual participants and disabling the identification of participants (given the data)—are extremely hard to reconcile even if data are managed by a single institution or study center. Here, data can be collected, then joined using participants' identifiers, and anonymized only before data analysis. This involves a high risk of leaking identifying information in case of security breaches before anonymization. In addition, anonymization needs to minimize the reidentification risk based on characteristic combinations of superficially anonymized data.

The problem is addressed by establishing *k-anonymity* [1], meaning that the identifying information of a person is indistinguishable from at least k−1 other datasets. Often k-anonymity can only be achieved by generalizing some attribute values, for example by reducing birthdates to ages. Further approaches to solve the reidentification problem involve a more sophisticated protection against attacks due to low data diversity and background knowledge in k-anonymized datasets [2] or small systematically randomized changes to the original data considered irrelevant for subsequent data analytics, but obfuscating individual characteristics [3].

Applying anonymization at the very end of data acquisition is most problematic in long-term studies while sensitive health data are stored nonanonymized over long periods of time, leaving the data vulnerable to attackers. With distributed data acquisition for the same patient, a downstream anonymization prior to data communication between trial centers becomes simply unfeasible, as identifying information needs to be shared to merge patient data.

Let's assume the following use-case as an example of potential risks. A clinical trial is performed to monitor population depression through weekly Web-based questionnaires. Each data entry (1 filling of the questionnaire) is stored as an individual record, and records are linked through the users' accounts.

Immediate anonymization is not an option here. Simple k-anonymization of the data would result in losing the possibility to link new records to existing datasets and, thereby, participant-individual trend assessment would not be possible. Instead, data might be pseudonymized. Pseudonymization removes all data directly identifying a person (such as name, address, and place and date of birth) and replaces this information with a generated data key, which, considered solely, will not unveil any hint leading to the real person, but is associated consistently with all data describing the same person. Consistent association of the same pseudonym with a person's data is the main task to be solved by pseudonymization approaches. Solutions established so far either (1) use a pseudonymization table or dictionary, which serves as a lookup device when new data need to be pseudonymized, or (2) adopt a function, deterministically calculating the pseudonym from identifying data [4] (ie, a hash function).

Even enhancing plain pseudonymization with encryption yields some problems. At any point in time, the records of the study participants could be grouped by pseudonym and possibly identifying information could be drawn from the combined information. Even the encryption of pseudonymized data by some master secret poses a risk, as an attacker would need to obtain only a single key.

The optimal solution to the attack scenario stated above would be to fully decouple identifying and study data during acquisition, meaning that no connection between the submitting participant and the records can be drawn at any point in time.

Such a solution needs to (1) remove all directly identifying data and (2) avoid using the same (pseudonym) key for datasets related to the same person. However, (3) records should still be grouped by study participant at the end of the study to analyze individual trends, and (4) participants should be enabled to trigger the deletion of all their data without disclosing their identity.

Objective

To the best of our knowledge, no such technique is yet available. Therefore, we propose a strong, decentralized pseudonymization technique based on shared public-private key cryptography that has the potential to de facto anonymize study data on acquisition while maintaining the possibility to group data by participant. In the following, we introduce the approach, demonstrate feasibility based on a proof-of-concept implementation, report related performance data, and address scalability based on adopting distributed computing (MapReduce).

Methods

Public-Private Key Scheme

To disconnect identifying participant information, such as user account name or email address, from study records, digital signatures based on asymmetric cryptography are employed. Users sign their records using a private key known only to them. These private keys are encrypted by a user's password and stored on a central server. Each time users add a new record to their dataset, they sign the record using their private key. Thus, each record is stored with a 1-time pseudonym (the signature), which can still be grouped based on available public keys without knowing the private key and, thus, without knowing the participant who created the record.

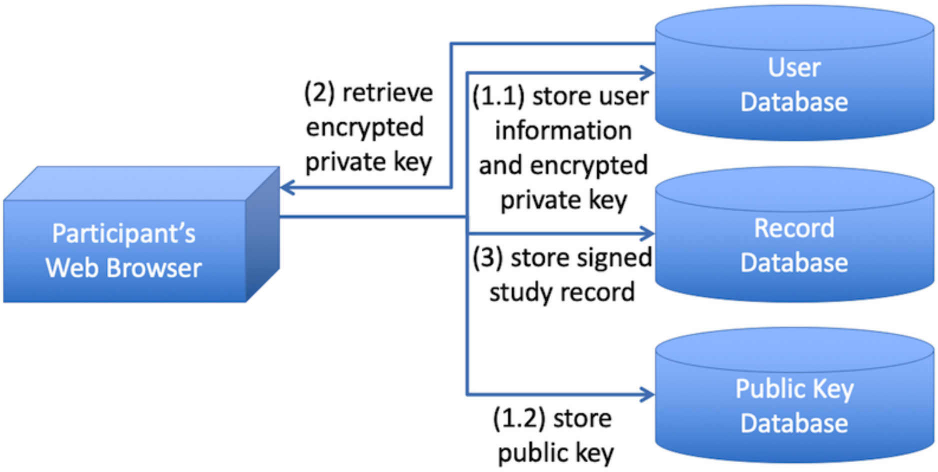
The matching public keys for each user are centrally stored on the server side. The records can be grouped by verifying the records' signatures using the combination of all public keys.

Wherever possible, cryptographic salt (a random sequence) should be added to hashes and signatures to increase entropy. The salt is stored alongside the hash or signature in the same database. For example, if a user submits the same answers to a questionnaire twice without salt, the same signature is created for both records; thus, they can be grouped without knowledge about the public key. With a random salt, 2 different signatures are created, since the sequences differ and grouping without knowledge of the public key is not possible.

System Setup

To ensure further privacy, the user data can be distributed over several servers (Figure 1). Identifying information such as email addresses or user names is stored in a *user database*. Study data such as filled questionnaires are stored in a *record database*, along with the record signature of the user and added cryptographic salt. For cryptographic data, a cryptographic hash with added salt of the user password is stored in the user database for identification; private keys are stored encrypted by the user's password in the user database with added cryptographic salt; and public keys are stored in a separate *public key database*.

Figure 1. Study database division and workflows.



Workflows

The main workflows can be split into 4 activities: (1) user registration, (2) user log-in, (3) record storing, and (4) record grouping.

During (1) *user registration*, a user with their respective public-private key pair is created. This is performed on the client side. The user chooses a username and password, and a public-private key pair is generated in the Web browser. The following information is then transmitted to the server infrastructure (Figure 1). First, the user database stores the potentially identifying user information such as user name and email address, as well as a cryptographic hash of the user's password with added cryptographic salt. In addition, the private key is symmetrically encrypted with the password as key and stored in the user database as well. Second, the public key database stores the public key. Both substeps are carried out independently and use separate databases.

On (2) *user log-in*, the user database is queried for the salt, and the hashed user password is generated on the client side and transmitted to the user database for authentication. If the authentication was successful, the encrypted private key is also returned and decrypted using the user's password in the Web browser.

Once logged in, the user can (3) *generate and store a new record*. The record is treated as binary data and hashed to reduce the amount of data to be signed. A cryptographic salt is added to the hash and the full sequence is signed with the private key. Record, salt, and signature are stored in the record database.

The records can be (4) *grouped* using the available public keys from the public key database. For this task, every record and salt from the record database is loaded and the signature is decrypted using each public key. If the decrypted value is identical to the hash of the record and salt, the signature was created with the corresponding private key (Figure 2). This is the well-known standard procedure for verifying a digitally signed document (Figure 3). The procedure is used here to assign all records that can be verified with the same public key to the same group. For the sake of a better visualization of our approach, we introduce a graphical shorthand notation of the verification process in Figure 4.

One crucial point should be highlighted: our approach never discloses the identity of the key owner. This is in sharp contrast to the usual verification of digitally signed documents, which relies on an approved association of the signer's identity with the public key. Instead, our approach collects the public keys of all participants in a public key store without any trace to the key owner's identity.

In order to group the records (ie, to join all records belonging to the same, but unknown, participant), all public keys are applied to all records. Records verified by the same public key must have been created using the same private key and, thus, belong to the same participant. Figure 5 shows the essence of this approach.

Figure 2. Grouping algorithm as pseudocode.

```
for each single_key in all_public_keys
{
    for each single_record in all_records
    {
        salted_signature = single_record.getSaltedSignature()
        serialized_document = single_record.getSerializedDocument()
        salt = single_record.getSalt()
        if decrypt( single_key, salted_signature ) =  hash(serialized_document + salt )
        {
            document_store( single_key ).append( document )
        }
    }
}
```

Figure 3. Standard procedure to verify a digitally signed document.

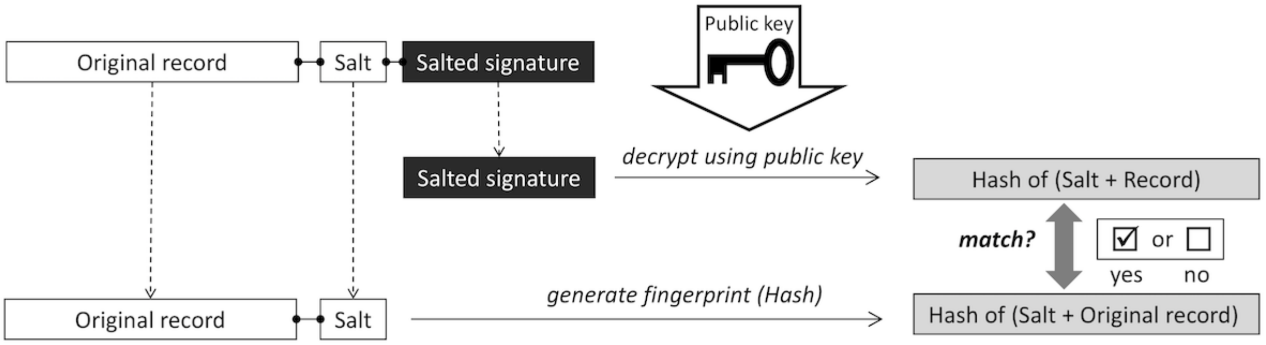
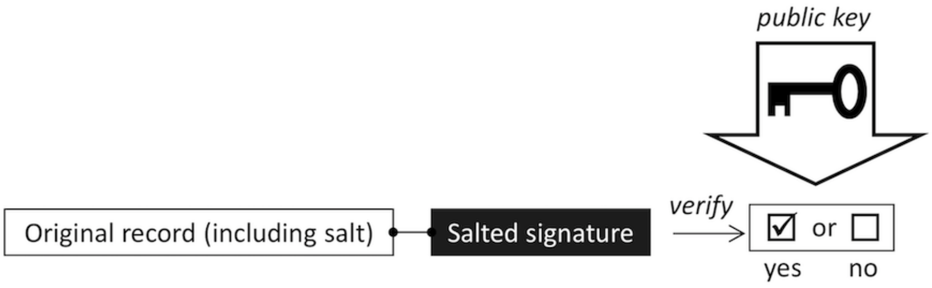


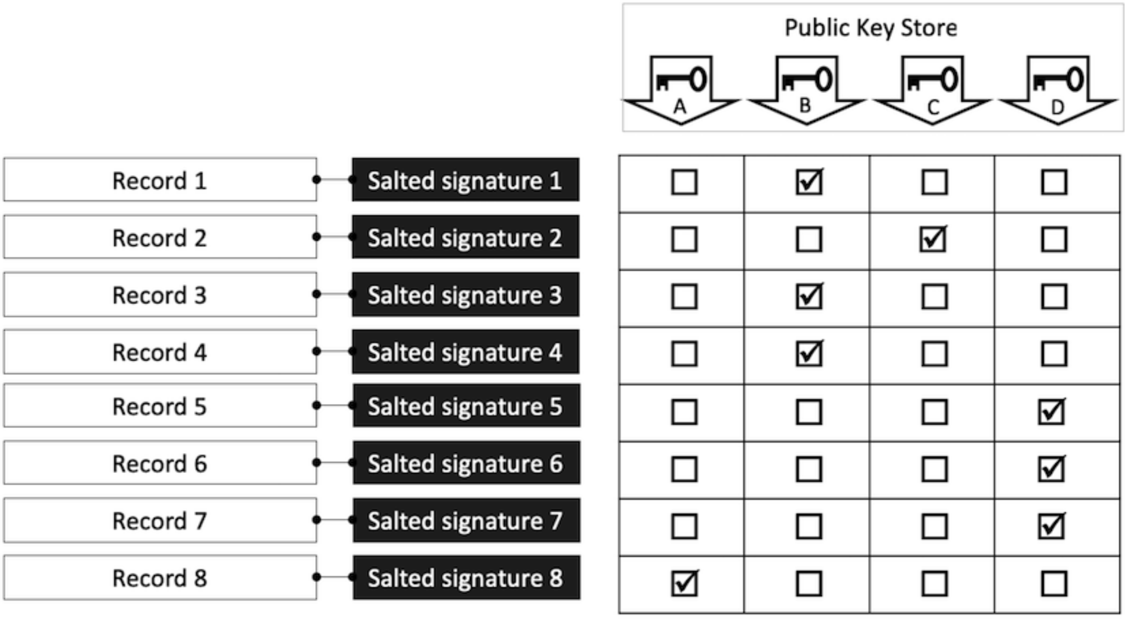
Figure 4. Shorthand graphical notation for the verification step of a digitally signed document as detailed in Figure 3.



Notification to Participants of Trial Analysis

In most trials, participants should be notified if specific diseases or other incidental findings are generated based on the collected data. In our example, a participant could be notified that depression has been diagnosed based on their replies. While our approach does not allow for direct communication to an individual patient identified based on their record, it still allows for directed feedback to just 1 distinct patient. This can be performed by encrypting a message with the public key associated with the diseased participant and sending this encrypted message to all participants (eg, through a mobile app or the Web interface of the trial). Since the message is encrypted with the distinct patient’s public key, only the associated private key is able to decrypt and read the message. Delivery to the correct patient is therefore guaranteed. However, since all patients need to be contacted for the right patient to log in, patients might be confused by potentially alarming messages. Thus, specific times should be scheduled for the communication (or log-in of the patient) or these findings should be communicated with general feedback from the study through the Web interface.

Figure 5. Record grouping using the public key store.



Public Key–Based Consent Management

While the approach actually relies on the mechanism of digital signatures, it is possible to implement privacy-preserving, participant-managed consent declaration. The participant's declaration of consent is treated exactly like a record. It is signed electronically by the participant's private key and enclosed in the record database. While the document is signed digitally, there is no need for further identifying information in the declaration. If all participants sign a declaration of consent, a study center will always be able to prove the righteous use of all personal data. The center only has to show that each record contained in the record database can be grouped with a consent declaration using the quasi-anonymous grouping mechanism introduced above (Figure 5).

Using a similar process, a participant can withdraw consent and trigger the deletion of his or her data. The system offers all participants the ability to sign and enter a delete statement. The delete statement is first treated like a usual record and grouped with all records related to the participant (including the consent declaration). Statements associated with a delete statement are flagged for deletion. A garbage collecting mechanism can then clean the record database. Additionally, for any group containing a delete statement and at least one record, the associated public key is also marked for deletion and then removed from the public key store. With these 2 steps, data of participants who withdraw consent are completely removed from the database (Figure 6).

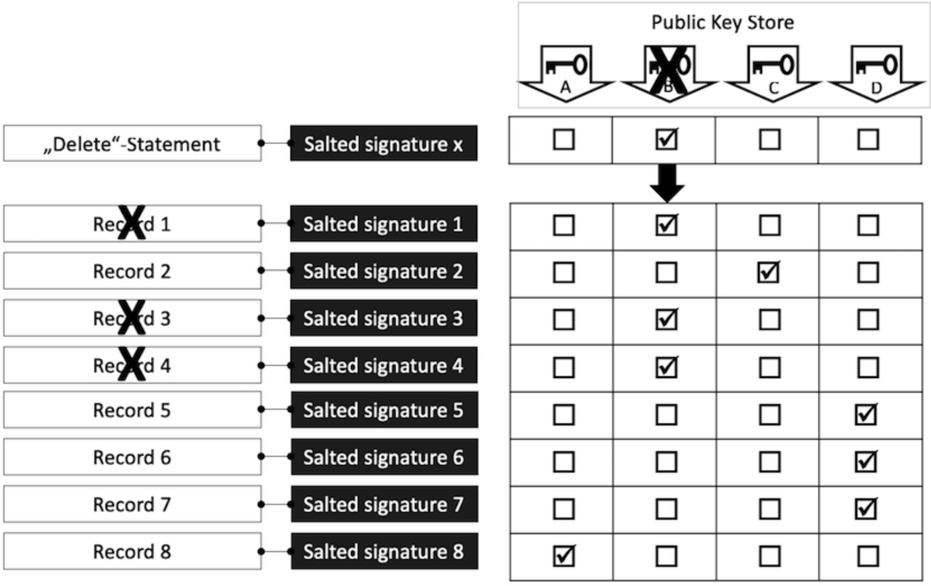
All operations for record grouping and consent management can be implemented as a MapReduce problem and can be addressed using big data technology such as Apache Spark or Hadoop.

Please note again that neither passwords nor private keys are transmitted or stored in clear text to the server at any point in time. Thus, records are grouped and consent is managed quasi-anonymously. In addition, no trusted third party is required for consent management (especially for handling revoked consent).

System Performance Evaluation

To evaluate the system regarding its computational performance and robustness, we performed several tests. The most expensive operation during user creation is the generation of asymmetric key pairs. Record generation and grouping speed is mostly dependent on the signing time of the record hash. Thus, we calculated runtimes for (1) user registration, (3) record storing, and (4) record grouping using different asymmetric encryption algorithms. In addition, we investigated and calculated possible collisions during calculation of hashes and public-private key pairs. Finally, we analyzed the potential threat of data exposure in various attacker scenarios.

Figure 6. Process for data management after withdrawal of consent.



Results

Runtime Considerations

We performed all tests using the Bouncy Castle application programming interfaces of the cryptographic algorithms [5]. We focus the report on the main runtime limitation of the asymmetric signing of documents. However, we tested different algorithms for hashing and symmetric encryption but do not report them separately. For the cryptographic hash function, we tested secure hash algorithm (SHA) -1 [6] and SHA-2 [7] with hash lengths of 256 and 512 bits. Since the performance impact on both record generation and the critical grouping operation was negligible, we chose SHA-256 as the hashing function.

We symmetrically encrypted users' private keys using the Advanced Encryption Standard (AES) [8]. We tested the standards AES-128 and AES-256 with the same result as for hash functions. The performance impact of the used block size did not affect the performance at all. We encrypted the users' private keys in the following simulations using AES-256.

We tested the following algorithms for the cryptographic signatures: ECGOST3410 (pure elliptic curve), elliptic curve version of the digital signature algorithm (DSA) (ECDSA) [9], SHA-256 DSA [10] (large integer factorization), and SHA-256 Rivest, Shamir, Adleman (RSA) [11] (large integer factorization). We tested all algorithms in a scenario with N=100 users and a mean of 1000 records with a size of 4096 bytes. We repeated each test 5 times and report the average result (Table 1).

Table 1. Average timings of common signature schemes for single operations using record length *l*=4096 bytes.

Signature scheme	Runtime (ms)		
	User registration	Record storing	Record grouping
ECGOST3410	0.350	0.342	0.198
SHA-256[a] DSA[b] 1024 bit	0.145	0.083	0.067
SHA-256 ECDSA[c]	0.057	0.089	0.011
SHA-256 RSA[d] 2048 bit	42.320	0.606	0.011
SHA-256 RSA 1024 bit	4.088	0.154	0.005

[a] SHA-256: secure hash algorithm with 256-bit hash length.

[b] DSA: digital signature algorithm.

[c] ECDSA: elliptic curve digital signature algorithm.

[d] RSA: Rivest, Shamir, Adleman.

Cryptographic Algorithm Collision Considerations

Two potential problems of the presented approach are (1) the chance of collision during signing and (2) the chance of collision during hashing. A collision in signing would yield an incorrect grouping of records. A collision in hashing would allow the injection of false data into the record database.

Given an existing record and its salted signature, a collision during hashing would enable an attacker to find a combination of different salt and record leading to the same hash. The attacker could then assign the preexisting signature to the new pair of salt and record and, thus, induce false but seemingly valid data. Since SHA-256 is used for hash computation, the hash length is 256 bits. The probability of a collision (P_{col}) can be approximated by applying the *birthday problem* [12] in the equation $P_{col}(m, o) = 1 – e [– m (m –1)/2^{o+1}]$, where m denotes the number of already computed hashes (record count m) and o is the output size of the hash function in bits. For the practical simulation, one therefore gets a collision probability of $P_{col}(105,256) \approx 0$. Actually, the probability of observing at least a single collision reaches 50% for $m \approx 4 \times 10^{38}$ hash operations. Thus, introducing false (and, more importantly, specific) data into the record database or having 2 records by the same person with the same hash is near impossible.

In the first case (signing), a collision can occur if 2 public keys are able to decrypt the same signature. Since the signing process is bijective, this means the 2 public keys have to be identical. However, the key length is even larger for the public keys than for the hashing algorithms; thus, the probability is even lower and poses no likely threat.

In contrast, the signing mechanism can almost guarantee the validity of the data, making forging (except for deletions) almost impossible.

Threat Analysis

Due to the layout of the system, the data are highly scattered between several (physically separate) database systems. We assume a retrospective attack after data have been entered for the following cases.

In the case where 1 database is compromised, that is an attacker gains access to any single database, it will not be possible to link a single participant to any of their study data, as none of the databases contains any connection between identifying data and records or between study and identifying data at the same time.

In the case where 2 databases are compromised, if the attacker gains access to the user and public key databases, they would not have access to study data. If the attacker gains access to the user and record databases, they would have no means of linking data from the 2 databases or even within the record database. If the attacker gains access to the record and public key databases, they would potentially be able to group records of the same user but would not have access to identifying information.

In the case where all 3 databases are compromised, the attacker would still only be able to group the records together if they know the cryptographic methods that were used. In this case, they would still have no connection between study data and identifying data, but they might be able to identify study participants based on behaviors or answers in the records.

In the worst case, to completely decrypt and link every single person with their study data (matching user name or email address with records), an attacker would have to gain access to each individual participant's computer and record the private key or password during the user log-in procedure, as well as gaining access to the records database and possibly the user database. This is not possible retrospectively, as the study is finished and no log-ins are made anymore.

Discussion

Principal Results

We present a cryptographic scheme for decentralized pseudonymization, participant information, and participant-managed consent declaration and withdrawal. The technical evaluation of runtime and collisions indicated not only that such a system is feasible, but also that runtimes are short enough to be integrated without notification or impairment of user experience. Especially when the focus is on small computational expenses on the user's end, the SHA-256 ECDSA is a good choice for signature generation, as it provides high security with a short runtime. SHA-256 ECDSA is also used in other cryptographic online tools such as Bitcoin. The chosen approach of signing each patient record individually also eliminates the risk of accidental or willful data tampering (except for record deletion).

This approach can be classified as decentralized pseudonymization, with the private key being part of the pseudonymization function. However, since the study personnel have no access to parts of the pseudonymization procedure and cannot interfere with it, the approach may potentially even be classified as de facto anonymization. If participants withdraw their consent, no depseudonymization is necessary. Records are marked for deletion and removed by a garbage collection mechanism without disclosing a participant's identity. Obviously, cooccurrence analysis of user interaction and deletion operations could leak identifying information. This risk can be minimized by (1) strictly separating the systems for user and record management, (2) cumulating record deletion requests for 2 or more users, and (3) avoiding a detailed log of user interactions.

The threat analysis showed that major effort would be needed to link identifying information with study data. Theoretically, gathering identifying personal information is only possible during the study and not after a study has been completed. A closed study might therefore be automatically anonymized. However, de facto security is highly dependent on implementation details, and—as history has taught us—even standard software libraries are prone to errors.

Limitations

The main limitation of the proposed method is that the user's private key becomes unrecoverable if the user forgets his or her encryption password. Without decrypting the private key, a user is not able to insert new records into the database to be grouped with older records. There are multiple possibilities available to reduce the risk of losing a key. One option would be to hand out the users' private keys in the form of smartcards. Losing a physical object is much less likely than forgetting a password. Another approach would be to store a copy of the private key encrypted with a secret obtained from the answers to a set of user-selected questions. Both approaches would decrease the risk of losing a private key but not totally eliminate it. In addition, these approaches might not be feasible in all settings or for all participants, which creates the risk of losing a study participant's data.

Similarly, since the trial is theoretically anonymized once the last participant has finished data entry, there is only an indirect way of communicating with study participants: through broad messaging to all participants. This might also not be feasible or successful in all cases. This limitation should be communicated to participants beforehand. A potential solution to this problem could be the addition of a trusted third party for those participants who value a direct follow-up more than anonymization of their data. Alternatively, an accompanying mobile phone app could hold the private key and could filter incoming messages and alert the user in case of information relevant to them.

Another limitation is possible collusion during cryptographic hashing. Our analysis showed that this is only a theoretical problem, but it may not be prevented completely.

Access control in general, and specifically the limitations of this approach, might discourage study participants from participating in a trial. However, accompanying mobile phone or tablet apps could reduce this burden by offering more convenient ways for password and key management (eg, face or fingerprint identifiers). However, this could in turn reduce security if private keys are stored nonencrypted on the device over longer periods of time.

Comparison With Prior Work

Prior work proposed the assignment of identical pseudonyms to records linked to the same participant or patient in order to enable record grouping or patient-specific data joins [4,13]. Pseudonymization is often combined with encryption to ensure both deidentification and confidentiality of the data [13]. Pseudonyms establish linked records within the database during the whole lifecycle of the database. Joining records by pseudonyms yields rich datasets per individual participant, which are, therefore, exposed to a greater risk of information-driven reidentification. In contrast, our approach avoids identical pseudonyms. Record groups are established on demand by the public key store. Thus, based on our approach, the public key store could be kept by a different organizational unit. Record grouping is then postponed to the time of data analysis, which reduces the reidentification risk during the data acquisition phase (especially the high risk due to low data diversity at the beginning of data acquisition; see Machanavajjhala et al [2]).

A main advantage of our approach is that no trusted third party is required for consent management. With respect to this point, our approach differs from an approach proposed by Aamot et al, which also adopted asymmetric encryption [14]: our approach completely avoids depseudonymization, but nonetheless enables withdrawal of consent and patient information.

Noumeir et al made the distinction between reversible and 1-way pseudonymization [15]. They argued that 1-way pseudonymization cannot support any notification of participants about incidental findings. They, therefore, proposed a symmetric encryption of data enabling reidentification by a trusted third party. Our approach is 1-way in the sense that a feedback on incidental findings cannot be propagated back to a distinct participant directly. However, through encryption and broad communication, a similar effect can be achieved. The only drawback is that it requires more activity by the user (ie, actively logging in to the system), which is not necessary in other cases.

Conclusions

We have proposed a novel cryptographic approach to pseudonymization that decentralizes the pseudonymization function and consent management in part to the study participants. Closed trials are thereby automatically anonymized, and a potential de facto anonymization at the study site during ongoing trials might be achieved. However, this claim requires further investigation and might be dependent on local privacy regulations. A prototypical implementation of the key cryptographic mechanisms of the trial software with grouping based on Apache Spark is available online [16].

Acknowledgments

We acknowledge the work of Matthias Wright on the prototypical implementation of the software.

Authors' Contributions

SJ wrote the manuscript, conceived the methodology, and conducted methodological analysis. SS conducted methodological analysis, performed experiments, and wrote and revised the manuscript. CS conceived the methodology, provided scientific oversight, and wrote and revised the manuscript.

Conflicts of Interest

None declared.

References

1. Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. Int J Uncertainty Fuzziness Knowl Based Syst 2002 Oct;10(05):571-588. [doi: 10.1142/S021848850200165X]
2. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. ACM Trans Knowl Discovery Data 2007 Mar 01;1(1):3-es. [doi: 10.1145/1217299.1217302]

3. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Halevi S, Rabin T, editors. Theory of Cryptography. Berlin, Germany: Springer; 2006:265-284.
4. Pommerening K, Reng M. Secondary use of the EHR via pseudonymisation. Stud Health Technol Inform 2004;103:441-446. [Medline: 15747953]
5. The Legion of the Bouncy Castle.: Legion of the Bouncy Castle Inc; 2013. URL: http://www.bouncycastle.org/docs/ [accessed 2019-01-02] [WebCite Cache ID 757aHkI7G]
6. Eastlake DE, Jones P. US secure hash algorithm 1 (SHA1). Reston, VA: The Internet Society; 2001 Sep. URL: https://tools.ietf.org/html/rfc3174 [accessed 2019-01-02] [WebCite Cache ID 757aGp6Pt]
7. Eastlake D, Hansen T. US secure hash algorithms (SHA and SHA-based HMAC and HKDF). Fremont, CA: Internet Engineering Task Force; 2011. URL: https://tools.ietf.org/html/rfc6234
8. Schaad J. Use of the Advanced Encryption Standard (AES) encryption algorithm in cryptographic message syntax (CMS). Reston, VA: The Internet Society; 2003 Jul. URL: https://tools.ietf.org/html/rfc3565.html [accessed 2019-01-02] [WebCite Cache ID 757aEOwhi]
9. Pornin T. Deterministic usage of the digital signature algorithm (DSA) and elliptic curve digital signature algorithm (ECDSA). Reston, VA: IETF Trust; 2013. URL: https://tools.ietf.org/html/rfc6979.html [accessed 2019-01-02] [WebCite Cache ID 757aD0IvS]
10. Turner S. Using SHA2 algorithms with cryptographic message syntax. Reston, VA: IETF Trust; 2010. URL: https://tools.ietf.org/html/rfc5754 [accessed 2019-01-02] [WebCite Cache ID 757a9yIst]
11. Jansen J. Use of SHA-2 algorithms with RSA in DNSKEY and RRSIG resource records for DNSSEC. Reston, VA: IETF Trust; 2009.
12. Wagner D. A Generalized Birthday Problem. In: Young M, editor. Advances in Cryptology - CRYPTO 2002: 22nd Annual International Cryptology Conference Santa Barbara, California, USA, August 18-22, 2002. Proceedings. Berlin, Germany: Springer-Verlag; 2002:288-304.
13. Heurix J, Neubauer T. Privacy-preserving storage and access of medical data through pseudonymization and encryption. In: Furnell S, Lambrinoudakis C, Pernul G, editors. Trust, Privacy and Security in Digital Business. Cham, Switzerland: Springer Nature Switzerland AG; 2011:186-197.
14. Aamot H, Kohl CD, Richter D, Knaup-Gregori P. Pseudonymization of patient identifiers for translational research. BMC Med Inform Decis Mak 2013 Jul 24;13:75. [doi: 10.1186/1472-6947-13-75] [Medline: 23883409]
15. Noumeir R, Lemay A, Lina J. Pseudonymization of radiology data for research purposes. J Digit Imaging 2007 Sep;20(3):284-295 [FREE Full text] [doi: 10.1007/s10278-006-1051-4] [Medline: 17191099]
16. Department of Medical Informatics, Division mHealth. Software. Aachen, Germany: Uniklinik RWTH Aachen, Department of Medical Informatics; 2019. URL: https://mhealth.imib.rwth-aachen.de/software/ [accessed 2019-02-01] [WebCite Cache ID 75rfqVu0A]

Abbreviations

AES: Advanced Encryption Standard
DSA: digital signature algorithm
ECDSA: elliptic curve digital signature algorithm
RSA: Rivest, Shamir, Adleman
SHA: secure hash algorithm

Edited by G Eysenbach; submitted 24.09.18; peer-reviewed by LA Lee, S Sadasivan; comments to author 28.11.18; revised version received 02.01.19; accepted 03.01.19; published 12.04.19.

Please cite as:
Jonas S, Siewert S, Spreckelsen C
Privacy-Preserving Record Grouping and Consent Management Based on a Public-Private Key Signature Scheme: Theoretical Analysis and Feasibility Study
J Med Internet Res 2019;21(4):e12300
URL: http://www.jmir.org/2019/4/e12300/
doi:10.2196/12300
PMID:30977738

Original Paper

Guideline Development for Technological Interventions for Children and Young People to Self-Manage Attention Deficit Hyperactivity Disorder: Realist Evaluation

Lauren Powell[1], BSc, MSc; Jack Parker[1], BSc, PhD; Val Harpin[2], MBBCHIR, MD, FRCP, FRCPCH; Susan Mawson[1], BSc, PhD, MCSP

[1]University of Sheffield, Sheffield, United Kingdom

[2]Sheffield Children's NHS Foundation Trust, Sheffield, United Kingdom

Corresponding Author:
Lauren Powell, BSc, MSc
University of Sheffield, ScHARR
30 Regent Street
Regent Court, Sheffield, S1 4DA
United Kingdom
Phone: 44 01142228272
Email: l.a.powell@sheffield.ac.uk

Abstract

Background: Attention deficit hyperactivity disorder (ADHD) is a complex neurodevelopmental disorder characterized by inattention, hyperactivity, and impulsivity. ADHD can affect the individual, the individual’s family, and the community. ADHD is managed using pharmacological and nonpharmacological treatments, which principally involves others helping children and young people (CAYP) manage their ADHD rather than learning self-management strategies themselves. Over recent years, technological developments have meant that technology has been harnessed to create interventions to facilitate the self-management of ADHD in CAYP. Despite a clear potential to improve the effectiveness and personalization of interventions, there are currently no guidelines based on existing evidence or theories to underpin the development of technologies that aim to help CAYP self-manage their ADHD.

Objective: The aim of this study was to create evidence-based guidelines with key stakeholders who will provide recommendations for the future development of technological interventions, which aim to specifically facilitate the self-management of ADHD.

Methods: A realist evaluation (RE) approach was adopted over 5 phases. Phase 1 involved identifying propositions (or hypotheses) outlining what could work for such an intervention. Phase 2 involved the identification of existing middle-range theories of behavior change to underpin the propositions. Phase 3 involved the identification and development of context mechanism outcome configurations (CMOCs), which essentially state which elements of the intervention could be affected by which contexts and what the outcome of these could be. Phase 4 involved the validation and refinement of the propositions from phase 1 via interviews with key stakeholders (CAYP with ADHD, their parents and specialist clinicians). Phase 5 involved using information gathered during phases 1 to 4 to develop the guidelines.

Results: A total of 6 specialist clinicians, 8 parents, and 7 CAYP were recruited to this study. Overall, 7 key themes were identified: (1) positive rewarding feedback, (2) downloadable gaming resources, (3) personalizable and adaptable components, (4) psychoeducation component, (5) integration of self-management strategies, (6) goal setting, and (7) context (environmental and personal). The identified mechanisms interacted with the variable contexts in which a complex technological intervention of this nature could be delivered.

Conclusions: Complex intervention development for complex populations such as CAYP with ADHD should adopt methods such as RE, to account for the context it is delivered in, and co-design, which involves developing the intervention in partnership with key stakeholders to increase the likelihood that the intervention will succeed. The development of the guidelines outlined in this paper could be used for the future development of technologies that aim to facilitate self-management in CAYP with ADHD.

(J Med Internet Res 2019;21(4):e12831) doi:10.2196/12831

KEYWORDS

attention deficit disorder with hyperactivity; technology

Introduction

Attention Deficit Hyperactivity Disorder, Prevalence, and Management

Attention deficit hyperactivity disorder (ADHD) is a highly comorbid [1] neurodevelopmental disorder, defined by 3 core symptoms: inattention, hyperactivity, and impulsivity. It has a worldwide prevalence of 3% to 5% in school-age children [2] and children and young people (CAYP) are most likely to be diagnosed with ADHD in the United Kingdom when they are at primary school [3]. This amounts to approximately 26 million children and adolescents, and this figure is rising globally [4]. Over the last 30 years, the number of people treated for ADHD in the United Kingdom has risen from 0.5 per 1000 to 30 per 1000 [3], and the annual health care costs for young people with ADHD in the United Kingdom are estimated at £670 million. CAYP with ADHD experience a number of ADHD-related difficulties including poor academic attainment, poor social relationships, increased likelihood of being suspended or expelled from school, and leaving school earlier than their peers [5]. In addition, genetic and contextual circumstances can also have an impact on the prevalence of the condition. ADHD is highly heritable [6], and those who are more socially disadvantaged are more likely to be diagnosed with ADHD [7,8]. Moreover, ADHD often continues to affect individuals into adult life [1,9].

ADHD management includes a combination of behavioral and pharmacological interventions [1]. There is strong evidence that pharmacological treatment and nonpharmacological interventions such as psychoeducation programs, behavioral interventions, and cognitive behavioral therapy have a major beneficial effect on the core symptoms of ADHD in approximately 80% of cases, at least in the short term [1,10]. ADHD can affect every aspect of an individual's life, and support from professionals and family members is limited. There is some evidence of short-term efficacy in managing the core ADHD symptoms, conduct disorders, social skills, self-efficacy, and emotional outcomes. However, CAYP often rely on clinicians and parents to manage the condition of members of the target population, and young people are often unwilling to engage in treatment [11], which limits ADHD self-management into adulthood [12]. Therefore, to attempt the prevention of the individuals falling into crisis when they reach adulthood, it is essential that the members of target population should learn how to self-manage their condition through co-designed interventions [13,14]. This includes exploring contemporary, innovative, and interactive methods of engaging CAYP with ADHD, such as the use of technology may improve their motivation and adherence to treatment. However, methodological limitations make it difficult to draw definitive conclusions from clinical trials [15].

Self-Management in Children and Young People With Attention Deficit Hyperactivity Disorder and Behavior Change Theories

People with long-term conditions (including ADHD) spend around 1% of their time interacting with a clinician, leaving 99% of their lives managing their condition themselves [16]. However, to self-manage a condition, behavior change is required. A number of theories have attempted to breakdown aspects of behavior that require change. For example, the Chronic Care Model (CCM) [17] identifies 6 elements [18] that are important factors for successful chronic care and prevention management that have previously been applied to the care of CAYP with ADHD [19,20]. These include and are not limited to the following:

1. The promotion of safe quality care; any self-management intervention for CAYP with ADHD will need to adhere to quality standards to ensure the content is reliable and appropriate.
2. Support should be based on evidence and what the patient's needs and preferences are; if the intervention does not adhere to what the patient wants or needs, the patient may be less likely to engage with it.
3. Self-management support should be provided to help patients manage their health and care; CAYP with ADHD should self-manage their condition to decrease the likelihood of them falling into crisis later in life.
4. Community resources should be available to improve access; resources should be available to facilitate and support the self-management of ADHD in CAYP.

Similarly, the Behavior Change Wheel (BCW) [21] provides a framework specifically for behavior change interventions and involves the *Capability Opportunity Motivation-Behavior* (COM-B) model, which refers to the interactions among "Capability," "Opportunity," "Motivation," and "Behavior." Capability refers to the psychological and physical ability to engage with an activity, opportunity refers to factors outside of the individual to ensure behavior change is possible, and motivation refers to brain processes that "energize and direct behavior." The COM-B model provides a useful framework of elements that influence behavior change and can indeed be applied to self-management. For example, to self-manage a condition, the individuals' behavior will need to change. To do this, they should be motivated and have the capability to change their behavior and be in the correct environment for the change to occur.

Furthermore, the Health Foundation states that people with long-term conditions need to have the knowledge, skills, and confidence to manage their condition "effectively in the context of...everyday life" [22]. These underlying principles of self-management and the principles from the CCM and the BCW are important for all long-term condition self-management, including ADHD in CAYP.

Technology Interventions for Attention Deficit Hyperactivity Disorder Self-Management in Children and Young People

Technology has been shown to have a large potential to improve the effectiveness and personalization of mental health interventions [15]. A number of attempts have been made to harness the technology to engage CAYP in self-managing their ADHD [23-41]. Examples include a handheld organizational device [37], computer games [25,27,34,39,41], programs [29,30,40], an augmented reality serious game [23], mobile apps to improve reading speed [33], executive functioning [35], and healthy sleep habits [38].

The results of these interventions have found an increased ability to remain on task at school [32], improved organizational skills [37], ADHD symptoms, and sleep [38]. It must also be noted that although a number of these studies have found positive results, it is unclear if these effects are maintained over a longer time period [24,31,33-36].

However, not all of these studies show positive or significant results. ADHD is a highly complex comorbid condition and it is therefore difficult to control for contextual differences using randomized controlled trial (RCT) methodologies. It is also possible that the uptake of each intervention among participants may vary [15]. Others may use the interventions in different contexts to one another with variable distractions [42,43].

Evaluating Complex Conditions

It is now understood that the steps taken for increasing evidence in complex conditions is no longer linear, and the updated Medical Research Council (MRC) Framework (2008) [44] places greater attention to the context in which interventions take place.

Figure 1 shows the outlining the Medical Research Council model of complex intervention development [44].

A total of 3 key components for the development of these complex interventions are outlined below:

1. Interventions should be clearly underpinned by existing theories. Theories that are based on existing knowledge can offer a clear way to underpin a rationale, which can assist with communication with stakeholders [45,46].
2. Interventions should be developed in partnership with key stakeholders [46-48].
3. Intervention developers should account for the context by which the intervention is developed by identifying what works for whom and under what circumstances. This means the intervention is more likely to be a success [42,43,49].

Other evaluation study designs such as RCTs and quasi-experimental studies only answer the question “What works?” and do no capture the complexity of complex conditions and interventions or the characteristics of the context in which the intervention is delivered [50]. This is important as the context, content, and outcomes of a complex intervention can involve a high degree of variance [50]. Therefore, if technological interventions are designed to be used with complex conditions such as ADHD in various contexts, it is imperative they are underpinned by theory and consider the contexts in which the intervention will be delivered [44]. A previous attempt to develop a complex intervention went beyond the question “What works?” and this involved a realist review that explored the question “what works for whom, under which circumstances and respects” [51]. However, to our knowledge, a realist evaluation (RE) has never been used to develop guidelines for the development of interventions .

Therefore, this study aimed to utilize an RE methodology [52] and involve key stakeholders (CAYP with ADHD, their parents and or carers, and specialist clinicians) in the development of theory- and evidence-based guidelines. The guidelines developed may help the future development of technological interventions that aim to help primary school–aged CAYP with ADHD self-manage their condition more effectively. Primary school–aged CAYP have been chosen as this is the most common age to be diagnosed with ADHD in the United Kingdom. RE aims to go beyond the “what works” question and answer the question “what works for whom, under which circumstances and respects.” RE also takes into account the complexities of the condition, the intervention, and the context by which it is delivered [52]. The use of underpinning behavior change middle-range theories (MRTs; see Table 1 for definition) will improve the generalizabiliy of the guidelines to more than 1 context. There is a need for these guidelines as existing frameworks are useful in terms of generalization to many conditions whereas CAYP with ADHD have complex needs that need addressing separately to ensure future interventions are suitable for them.

Figure 1. Outlining the Medical Research Council model of complex intervention development.

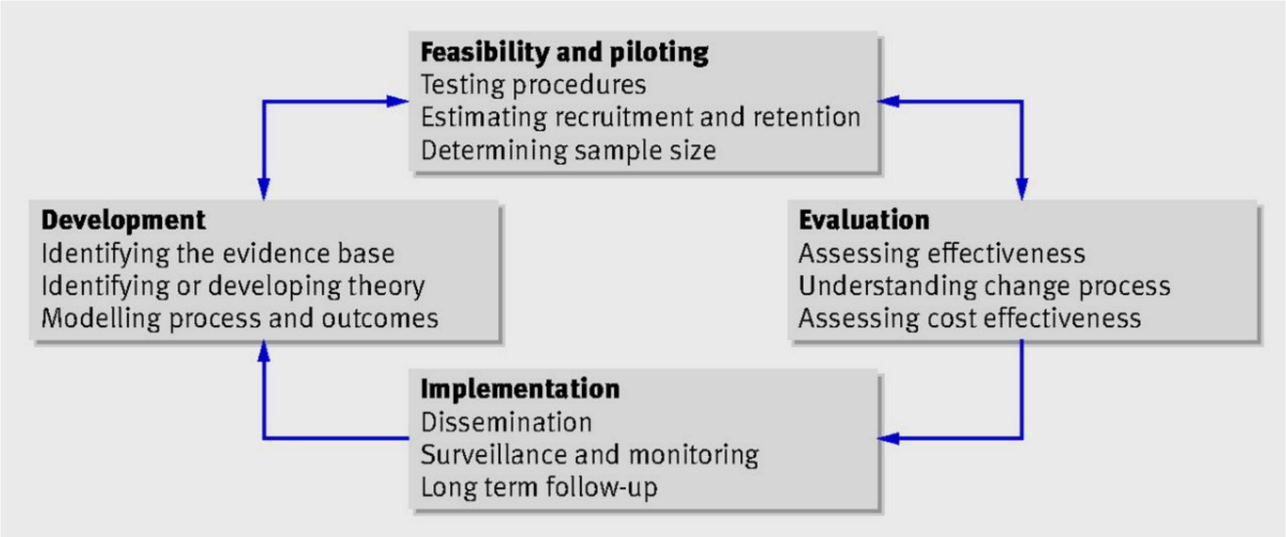


Table 1. Definitions of context, mechanism, and outcomes.

Term	Definition
MRT[a]	A theory that can be used to explain specific parts of an intervention is called an MRT. MRTs are identified at the beginning of this process and examined throughout the process and for this study, during data collection.
Context	The environment or "backdrop" of an intervention is called Context. Context can change over time, which could reflect aspects of change while an intervention is implemented [54]. The context may limit or allow the mechanisms.
Mechanism	This refers to aspects ("resources") that are a result of the intervention and the response to those resources, for example, cognitive, motivational, and emotional [54].
Outcome(s)	Outcomes (intended or unintended) refer to what may happen because of an intervention. For example, variable context may create an unintended outcome, which could be vital to intervention delivery.

[a]MRT: middle-range theories.

Methods

Principles of Realist Evaluation

RE has been shown as an effective framework for evaluating complex health interventions [43]. The aim of RE is to explore how a mechanism may cause a different outcome when in different contexts (see Table 1 for definitions) [52]. The process adopted for this study is outlined in Figure 2. The RE approach outlined in this study has been guided by Realist And Meta-narrative Evidence Synthesis: Evolving Standards II reporting standards for RE [53] and has been followed by the process stipulated in Pawson et al’s study, 1997 [52].

The Five Stages of This Study

The 5 stages of this study are as described in the following sections:

Stage 1: Identifying Propositions

Propositions are comparable with that of hypotheses that predict what is believed to occur in a given situation or within research. Developing the propositions for this study involved authors LP and JP exploring theoretical concepts from the literature that derives from behavior change and human-computer interaction theories (see Table 2) that could underpin a technological intervention that aims to help CAYP with ADHD self-manage their condition. Agreement of these concepts was reached through discussion among all the authors. The product of Stage 1 was a list of propositions.

Stage 2: Identifying a Theoretical Framework

Using the principles of RE [43], a theoretical framework was formed to underpin the development of the intervention guidelines, that is, concepts within identified theories could underpin specific components (or “mechanisms”) of an intervention. The theoretical framework was based on theories that can be applied to educating CAYP with ADHD and human computer-interaction (see Table 2), and it was constructed by authors LP and JP.

Stage 3: Context Mechanism Outcome Configuration Generation

After the propositions (Stage 1) and the theoretical framework (Stage 2) were developed, they were set out as context mechanism outcome configurations (CMOCs) during Stage 3 of this process. Authors LP and JP generated the CMOCs. Table 3 outlines some examples of the CMOCs generated during Stage 3.

Figure 2. Outlining the process of generating, validating and refining propositions and context mechanism outcome configurations. This process lasted between May and September 2018.

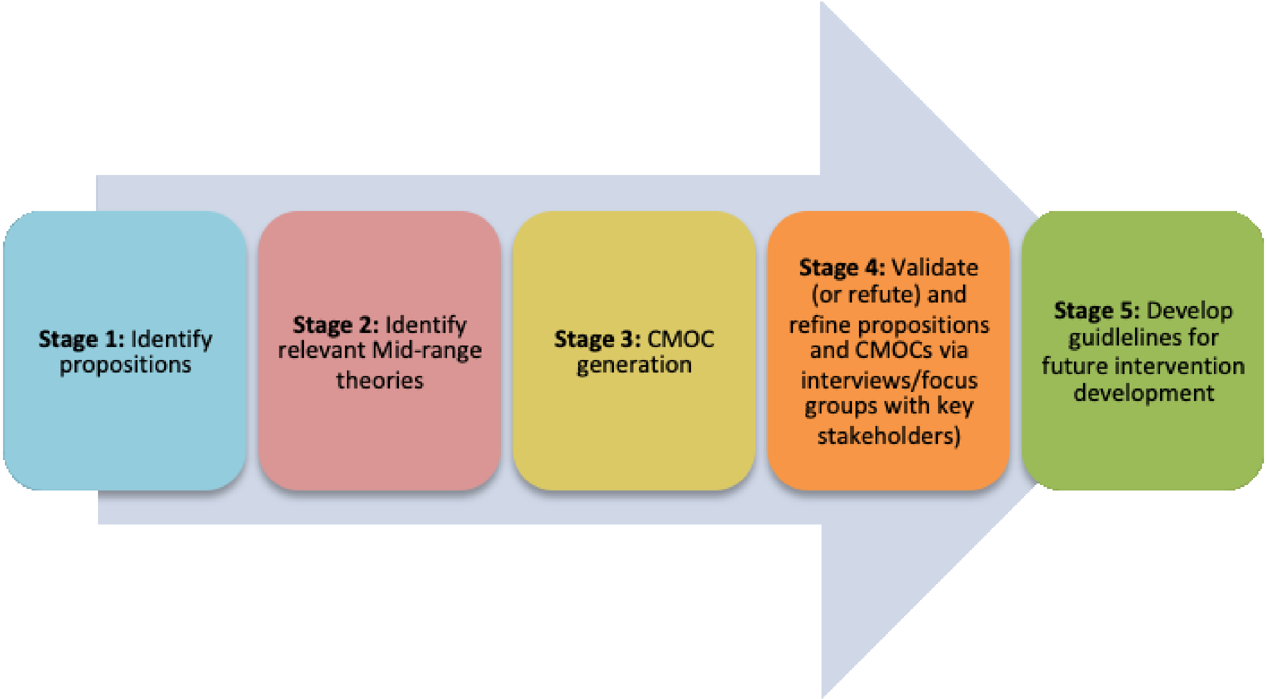


Table 2. Product of Stage 2: demonstrating how middle-range theories underpin the intervention guidelines.

Middle-range theories	Ingredients and middle-range theory link	How intervention could incorporate the ingredients
CC[a], OC[b], OST[c], SLT[d], SRT[e], ED[f], DDT[g], ELT[h], SCT[i], SDT[j], OIT[k], BCW[l], CCM[m]	Reward (OC, CC, DD, SDT, BCW)	Immediate rewards for all correct responses to engage and motivate the user.
	Stimulation (OST, ED)	User can move on to different available sections of the intervention and previous work will be saved to return to later. User has the choice to carry out intervention activities electronically or on paper.
	Sequential learning (ED)	All "sections" of intervention to not be available at once (preventing overstimulation). Different sections become "unlocked" once other sections are completed.
	Self-efficacy (SLT and SCT)	Intervention will provide the users with the opportunity to self-evaluate their performance, by receiving feedback from the intervention (eg, stars and coins) and from others (verbal persuasion or encouragement).
	Learning (ELT)	Paper-based activities will be available for those with limited access to a device (eg, sharing with siblings or limited device access at bedtime) and/or internet.
	Independent practice (SLT)	Used in the absence of a clinician.
	Social regulation (SRT and CCM)	Section that teaches user techniques to self-manage ADHD[n], for example, anger management.
	Social Learning (SLT)	Intervention should provide scenarios of social situations where the user can make appropriate decisions (reinforced with immediate rewards).
	Social cognition (SCT)	Setting short-term, meaningful, and relevant goals for the users to motivate them to engage with the intervention.
CD[o], UID[p], and CCM	Stakeholder involvement in design (CD, UID, and CCM)	Stakeholders should be involved in the design and development of the intervention to increase intervention success.
CC, OC, OST, SLT, SRT, ED, DDT, and ELT.	Self-monitoring	Users monitor their performance independently.
	Reinforcement	Intervention should provide positive feedback where applicable and they can share this with others.
	Self-management	Intervention should give the users opportunities to problem solve, make decisions, and take action in real life scenarios based on what they have learned.
	ADHD Knowledge and understanding	Intervention should provide the users with accessible information to help them better understand ADHD so they can more optimally self-manage it.

[a] CC: Classical Conditioning [55].

[b] OC: Operant Conditioning [56].

[c] OST: Optimal Stimulation Theory [57].

[d] SLT: Social Learning Theory [58].

[e] SRT: Social Regulation Theory [59].

[f] ED: Executive Dysfunction [60].

[g] DDT: Dynamic Developmental Theory [61].

[h] ELT: Experiential Learning Theory [62].

[i] SCT: Social Cognitive Theory [63].

[j] SDT: Self Determination Theory [64].

[k] OIT: Organismic Integration Theory.

[l] BCW: Behavior Change Wheel [21].

[m] CCM: Chronic Care Model [17].

[n] ADHD: attention deficit hyperactivity disorder.

[o] CD: Co-design [13,14].

[p] UID: user interface design.

Table 3. Product of Stage 3: context mechanism outcome configuration examples.

CMOCs[a]	Plausible mechanism: "What"	Contexts: "for whom" and "in what circumstances"	Possible outcomes
CMOC 1	Receiving feedback from the intervention might improve the users' confidence by confirming performance.	Internet and intervention accessible at home, used independently of clinician.	Development of self-efficacy
CMOC 2	Positive reinforcement (reward) may motivate the user to use the intervention.	Intervention should give positive rewarding feedback to the user.	Increased understanding of condition and self-management

[a]CMOC: context mechanism outcome configuration.

Stage 4: Validation and Refinement of New and Existing Context Mechanism Outcome Configurations

CMOCs were then validated and refined by conducting interviews with CAYP with ADHD, their parents/carers, and specialist clinicians. Author LP conducted the interviews and they were conducted at the participant's convenience. Clinician interviews were undertaken at the clinicians' workplace and young persons' and parents' interviews took place in their homes.

Participants

Participants were recruited to adhere to the sampling frame below.

- CAYP with ADHD and their parents/carers
  - Males and females
  - CAYP with autism spectrum disorder (ASD) and without ASD
  - Families who live in the 10% of most and least deprived areas of the United Kingdom [65]
  - CAYP with ADHD aged 8 to 11 years
- Clinicians
  - A sample that includes ADHD specialist nurses, a pediatrician, and a psychiatrist.
  - Clinicians who work at Child and Adolescent Mental Health services and pediatric neurodisability services.

Recruitment

CAYP with ADHD and parents/carers were recruited via a database held by the research team. Clinicians were recruited via the National Health Service (NHS) in the South Yorkshire region. Participants were recruited until data saturation was achieved [66]. The eligibility criteria for CAYP with ADHD were (1) aged 8 to 11 years and (2) diagnosed with ADHD. Parents/carer (1) must have been a parent/carer of a young person with a confirmed ADHD diagnoses and (2) must have been able to provide details of the ADHD medication the young person was prescribed. Clinicians had to be employed by a service that treats CAYP with ADHD and has experience of working with this population.

Procedure

Semistructured interviews focused on initial propositions that were tested and refined. CAYP with ADHD, their parents/carers, and clinicians provided interview data to test the propositions. The study received ethical approval from the University of Sheffield's School of Health and Related Research Ethics Committee (Ref: 021203) and received NHS Health Research Authority and Research and Development local approval. Interviews took place in the CAYP/parents/carers' homes and clinicians' workplaces. All participants provided written informed consent or assent (CAYP only).

Participants were asked (age appropriate) questions that were derived from the propositions. Questions included the following:

- What type of feedback do you think your child would like and why? (parent/carer)
- What do you think the role of friends and family could be for supporting CAYP with ADHD with a technological intervention? (clinicians/parent/carer)
- If you play a computer game, do you like to collect things like coins, stars, points? (CAYP with ADHD)

Parents/carers provided ADHD medication details for their child (where applicable) and completed a Swanson Nolan and Pelham IV questionnaire to provide a measure of their child's current ADHD symptoms.

Data Analysis

Analysis focused on refining and generating new CMOCs. Principles of thematic and framework analysis were adopted [67,68]. Guidelines were identified on the basis of existing CMOCs (framework analysis approach), and when data did not fit with existing CMOCs, new CMOCs were generated (thematic analysis approach) [43].

Stage 5: Development of Guidelines

This was based on the refined and newly generated CMOCs from Stage 4. The final guidelines aim to provide a set of recommendations for designing a complex technological intervention that aims to help CAYP with ADHD self-manage their condition. The guidelines also provide advice regarding the environment in which the intervention should be delivered. The CMOCs refined during Stage 4 were used to form the content of the guidelines. The guidelines can be found in Multimedia Appendix 1. Author LP initially put the guidelines together and then discussed the guidelines with the rest of the research team (authors JP, VH, and SM) and refined them accordingly.

Results

Participant Characteristics

A total of 21 participants (7 CAYP, 8 parents, and 6 clinicians) were recruited from July 2018 to October 2018. Participant demographic information is included in Table 4 (CAYP), Table 5 (parents), and Table 6 (clinicians). All parents were able to provide information regarding their child's ADHD medication. All interviews were transcribed verbatim. During analysis, agreement between the 2 primary coders was high.

Initial Propositions (Stage 1)

Overall, 9 propositions were identified by author LP and checked for accuracy by author JP. They were then tested against the interview data and refined:

1. If the user receives feedback from the intervention, then the user's confidence may be improved.
2. If the user can access downloadable resources from the intervention, then the user may generate a deeper understanding of the concepts covered within the intervention.
3. If the users can choose personalizable characters and a variety of modules within the intervention, then this may enable them to maintain stimulation to carry out the task.
4. If the users receive positive reinforcement (reward) from the intervention, then this may motivate them to use the intervention.
5. If the users engage with social scenarios within the intervention, then they may make more appropriate social decisions in the future, which may help enhance social relationships.
6. If the users engage with the intervention, then they may gain a better understanding of their ADHD.
7. If the users engage with the intervention, then improved self-management of their ADHD may improve relationships with friends and family.
8. If the user gains encouragement from friends/relatives to use the intervention, then this could reinforce the user's engagement with the intervention.
9. If short-term meaningful goals are set for the users via the intervention, then this could encourage them to engage with the intervention.

Table 4. Demographic information of children and young people with attention deficit hyperactivity disorder.

Study ID	Gender	Age (years)	Other diagnosis	ADHD[a] medication	Medicated during interview?	SDI[b]	Inattention SNAP[c] Score	Hyperactivity or Impulsivity SNAP Score	Connors Index	Combined SNAP Score[d]
YP1	Female	11	ASD[e]	Concerta	Yes	820	2	1.66	1.8	1.82
YP2	Male	9	N/A	N/A	No	13513	2.55	2.89	2.5	2.65
YP3	Female	8	ASD	Usually 27 mg Delmosart	No	17403	3	3	2.7	2.9
YP4	Male	10	N/A	Delmosart 36 mg	Yes	23954	1.78	2	2.6	2.13
YP5	Male	11	N/A	Delmosart 36 mg+27 mg	Yes	4913	1.56[f]	2.22	2.4	2.06
YP6	Male	9	N/A	Delmosart 36 mg	Yes	1318	1.78	2.56	2.2	2.18
YP7	Male	8	Attachment disorder	Elvanse, 40 mg	Yes	32596	1.67	1.88	1.9	1.82

[a] ADHD: attention deficit hyperactivity disorder.

[b] SDI: Social Deprivation Index. 1 is indicative of the most deprived area in the United Kingdom and 32844 is the most affluent area in the United Kingdom.

[c] SNAP: Swanson, Nolan, and Pelham Questionnaire. SNAP Scores: Scores indicative of ADHD are as follows: Inattention: 1.78 and above; Hyperactivity/Impulsivity: 1.44 and above; Connors Index: 1.67 and above; Combined score: 1.63 and above.

[d] Average score across Inattention, Hyperactivity/Impulsivity and Connors Index subsections.

[e] ASD: autism spectrum disorder.

[f] Please note YP6 does not meet the threshold for one SNAP component. They did meet the criteria for all other SNAP domains.

Table 5. Demographic information of parents of children and young people with attention deficit hyperactivity disorder.

Participant ID	Comorbid condition of child	Age of child (years)
P1	ASD[a]	11
P2	—[b]	9
P3[c]	ASD	8
P4	—	10
P5	—	11
P6	—	9
P7	Attachment disorder	8
P8[c]	ASD	8

[a] ASD: autism spectrum disorder.

[b] Not applicable.

[c] P3 and P8 are the parents of the same child and were interviewed together.

Table 6. Demographic information of clinicians demonstrating 8 months to 18.5 years of experience of working with children and young people with attention deficit hyperactivity disorder with a mean of a total of 6.9 years of experience.

Participant ID	Gender	Job title	Clinical experience with children and young people with attention deficit hyperactivity disorder (years, months)
C1	Male	Registrar psychiatrist	2 years
C2	Female	Consultant pediatrician	4 years 6 months
C3	Female	Nurse clinical specialist	18 years 6 months
C4	Female	Nurse prescriber	8 months
C5	Female	Consultant community pediatrician	10 years
C6	Female	Consultant community pediatrician	6 years

Testing the Propositions

Overall, 7 themes were identified: (1) positive rewarding feedback, (2) downloadable gaming resources, (3) personalizable and adaptable components, (4) psychoeducation component, (5) integration of self-management strategies, (6) goal setting, and (7) context (personal and environmental). These themes focused on testing the 9 initial propositions.

Positive Rewarding Feedback (Propositions 1 and 4)

All participants expressed a wish for immediate positive reward when the user may select a correct response. Of the participants, 1 said that when he or she gets a reward, for example, a sticker at school, it makes him or her feel “proud” (YP5). Examples of instant reward could be auditory confirmation of a correct response and collecting items such as coins, stars, diamonds or trophies. The reward (and the intervention itself) should also be visually attractive:

> *I think that [instant positive reward] will really help his self-confidence.* [P4]

Another clinician states the following:

> *I think a lot of the games nowadays build up points and it makes sense...having reward builds up their self-esteem.... And just makes them feel happier.* [C4]

In addition, all 21 participants suggested that the instant positive reward component would motivate the user to engage with the intervention. They also felt that additional motivation to engage with the intervention could involve personalizing the reward (n=12), that is, the users can choose their rewards (eg, coins, trophies, stars) because of the following reason when referring to CAYP with ADHD:

> *tend to get bored quite quickly.* [C2]

A total of 11 participants stated reward could also be given by providing different levels where the use could “level up” or open “new areas” once a previous level is “completed.” Most of the CAYP (n=5) and 2 parents wanted these levels to increase in difficulty:

> *I like harder and harder cos if you do harder and harder you get better and better at it.* [YP7]

However, 2 CAYP (YP3 and YP6), 1 parent (P7), and 2 clinicians believed that if levels were too challenging for the users, it could cause frustration and demotivate their engagement with the intervention. Therefore, 1 clinician suggested that there could still be levels and areas to create choice, allowing the users to feel they are progressing, but these levels could have an option to make them easier:

> *...simplify the challenge so you could make the challenges harder...but there could be a simplify option that the kids could use and so the kids that do get frustrated can simplify it and get it done.* [C1]

Downloadable Gaming Resources (Proposition 2)

If applicable, the option of using downloadable resources could be made available for when the intervention may not be accessible, for example, if the child/young person has to share a device with siblings, has limited screen time (eg, before bed time), or is away from home (eg, in the car or on holiday). Participants wanted downloadable resources to have a gaming component, including quizzes, mazes, word searches, crosswords, coloring in, or origami activities (5 CAYP, 3 parents, and 2 clinicians).

> *Including quizzes cos I like quizzes.* [YP1]

Personalizable and Adaptable Components (Proposition 3)

A total of 5 CAYP, 4 parents, and all 6 clinicians requested that the technology should be personalizable and include adaptable avatars, that is, characters they can personalize by changing hair/eye color, gender, clothing, and skin color that can be adapted as and when they wish. Moreover, 4 clinicians believed this was so that the user could "relate" to the intervention and its content:

> *That it's [the language] not too clinical and that they can actually relate to it... It's the relating to it really that's most important. ...you have to be really careful that its not so generalised that they can't relate to it.* [C3]

Moreover, 1 parent (P4), 1 young person (YP4), and 1 clinician (C4) emphasized the importance of having the correct amount of stimulation to ensure the users are not over or under stimulated:

> *You don't want to over-stimulate them, but you want them to have that draw, I think its finding the right balance between overload and sort of retaining err concentration.* [P4]

Psychoeducation Component (Proposition 6)

A total of 5 CAYP, 5 parents, and all 6 clinicians believed it was an important aim for the users to have a good understanding of their ADHD. It was also considered important by 2 clinicians (C2 and C3) that the positive aspects of ADHD should be highlighted through examples of others who have ADHD and have been successful, such as celebrities, as they believed there was a lot of negativity surrounding the condition. Moreover, 1 clinician believed that it could be "life changing" (C1).

Another clinician stated the following:

> *Because I want to know about ADHD, what it does and what it effects in your body.* [YP5]

Moreover, 1 parent stated the following:

> *Knowledge is power and just giving her the confidence, increased self-esteem.* [P3]

A clinician stated the following:

> *I think it could be massive for them across the board it could help them at school, help them learn, help them make friends, help with their relationships with others...* [C1]

A total of 5 clinicians wanted interventions for CAYP to self-manage their ADHD to be more positive while not "glossing over" some of the difficulties. All 6 clinicians expanded on the above and stated that when CAYP with ADHD act incorrectly or impulsively, they often feel bad about themselves, and having knowledge about their condition could help prevent this. Overall, 3 CAYP, 4 parents, and 4 clinicians emphasized the importance of understanding the users' ADHD so the users can explain it to their friends:

> *Cos if I know more I can tell people more about the like what I've got [ADHD] so they know what it means.* [YP4]

Integration of Self-Management Strategies (Propositions 5 and 7)

Overall, 5 CAYP, 4 parents, and all 6 clinicians believed an intervention should include strategies to help the children self-manage their ADHD, such as anger management strategies. Moreover, 1 participant stated that he counts to 40 for a total of 3 times to calm down (YP7).

Another parent stated the following:

> *He can learn sort of techniques you know sort of self-management techniques trying to calm himself down.* [P4]

Another self-management strategy discussed was animated "social scenarios" with alternate endings for the users to choose from to help them understand what acceptable behavior is and is not in social situations. This idea was favored by clinicians (n=5), CAYP with ADHD (n=3), and parents (n=3). Overall, 2 clinicians (C5 and C6) stated this could be beneficial because similar "social stories" are already used with CAYP with ADSD, which is comorbid in many CAYP with ADHD:

> *She seems to learn a lot through like watching videos... if she wants to know how to do something, she goes on YouTube.* [P1]

A clinician stated the following:

> *I really like the idea of scenario-based teaching.* [C2]

Another clinician said:

> *[Social scenarios] sound like a similar principle to the social stories we use with the children with autism we see. I think that could be useful as it could help the children to reflect on what they might do in a situation before they are in the heat of the moment.* [C5]

Goal Setting (Proposition 9)

Overall, 6 parents liked the idea of goal setting within a technological intervention. They liked the idea of short-term goals because of poor working memory in CAYP with ADHD, which means they may find it challenging to process longer-term goals.

Context (Propositions 6 and 8)

It was found that the variable context an intervention is delivered in could affect the outcome it may have, and these contexts could be divided into environmental and personal.

Table 7. Refined context mechanism outcome configuration examples that support initial propositions.

CMOCs[a]	Plausible mechanism: "What"	Contexts: "for whom" and "in what circumstances"
CMOC 1	Receiving positive rewarding feedback from the intervention might improve the users' confidence by confirming performance.	Internet and intervention accessible at home and used independently of clinician. Intervention should be colorful and not too text heavy.
CMOC 3	Enabling the user to choose personalizable and adaptable characters of majority and minority groups and a limited number of "modules" will maintain stimulation to carry out the task.	The intervention will give positive and rewarding feedback to the user. Users will also have their own user area so that they can return to previous work and carry on where they left off.

[a]CMOC: context mechanism outcome configuration.

Table 8. Additional context mechanism outcome configurations generated from context mechanism outcome configuration validation with key stakeholders.

CMOCs[a]	Plausible mechanism: "What"	Contexts: "for whom" and "in what circumstances"
CMOC 10	Users will have a better understanding of their ADHD so they can explain it to others (friends/family).	The intervention will provide age-appropriate information to improve the users' knowledge and understanding of their ADHD[b] and provide suggestions on how to explain their ADHD to others.
CMOC 11	An indication of improvement or progress such as leveling up will motivate adherence.	The intervention will provide the user with varying game levels to keep them engaged and motivated to use the intervention. A "simplify option" will also be available to keep frustration levels down where applicable.

[a]CMOC: context mechanism outcome configuration.

[b]ADHD: attention deficit hyperactivity disorder.

Personal Contexts

Overall, 1 parent (P3) and 1 clinician (C5) stated that some CAYP with ADHD also have dyslexia and may struggle to read text; therefore, the background color to any included text should be adaptable. This is because some people with dyslexia find it easier to read text on specific background colors. This could enable the user to access the information more easily. Overall, 1 parent (P1) and all 6 clinicians also believed it is important that the information presented should be developmental and age-appropriate and the language should be suitable to ensure the user can understand the material provided:

> *A whole variety of those different [background] colours then that would make it much more accessible. It would make it easier for them [CAYP with dyslexia as well as ADHD] to read, it could stop the words and the letters moving, it makes it so they can actually read what's written rather than it being a sea of text they can't access. There's a huge overlap between lots of condition like dyspraxia, dyslexia, ADHD, Autism.* [C5]

Environmental Contexts

Overall, 13 participants (3 CAYP, 4 parents, and 6 clinicians) believed they would be more motivated to engage with a technological intervention if they had encouragement and support from close friends or relatives. Moreover, 1 young person (YP1) stated that her family and her dog could get in the way if she was to use an intervention of this nature, which could affect the effect the outcome intervention has on the user:

> *I think it will be good for them to do on their own but I think it will be good for other people to know what they have looked at so they can reinforce if they have any questions.* [C3]

A total of 6 parents believed that supporting their child with an intervention that helps them self-manage their ADHD could help build their relationship with their child.

Stage 4: Context Mechanism Outcome Configuration Refinement

As a result of validating CMOCs with key stakeholders, existing CMOCs have been refined (see Table 7 for examples) and 2 more CMOCs have been developed (Table 8). All CMOCs can be found in Multimedia Appendix 2.

Discussion

Principal Findings

This study aimed to present an RE approach to develop guidelines that may help the future development of technological interventions, which aim to help CAYP with ADHD self-manage their condition more effectively. A total of 7 key themes emerged from the interviews with key stakeholders: (1) positive rewarding feedback, (2) downloadable gaming resources, (3) personalizable and adaptable components, (4) psychoeducation component, (5) integration of self-management strategies, (6) goal setting, and (7) context (environmental and personal). Importance was placed on the variable environmental and personal context in which such an intervention could be delivered; importance was additionally placed on how these contexts could affect the outcomes of the interventions.

Comparison With Previous Work

Positive Rewarding Feedback

All participants identified the need for an instant positive reward within a technological intervention for CAYP with ADHD. This is supported by behavior change MRTs such as classical conditioning, which states that unconscious behavior will change when a stimulus is repeatedly paired with a particular response such as rewards [55]. Similarly, Operant Conditioning is when an individual repeatedly makes an association with a stimulus, such as reward or punishment [56]. These theories explain why the administration of reward can change behavior. Dynamic developmental theory states CAYP with ADHD have a shorter “window” between behavior and a reward response for them to make the association between the behavior and the positive response [61]. This explains why the reward should be immediate. Bandura’s theory of self-efficacy also states that gaining confidence by achieving and accomplishing a task can increase an individual’s self-efficacy. This is referred to as “Mastery Experiences” [69]. The BCW states that the individual needs to be motivated for behavior to change [21], and reward could motivate a child with ADHD to engage with an intervention.

Downloadable Gaming Resources

Some participants (5 CAYP, 3 parents, and 2 clinicians) liked the idea of having the option of being able to print off resources that complement the technological intervention in the event that technology is not available (eg, before bedtime and away from the home). This could be important as CAYP with ADHD are overrepresented in socially deprived areas [7,8] and may not have access to technology. It would also provide the user the opportunity to have an experience away from a screen and could help supplement learning by conducting a physical action. The latter claim is supported by John Dewey’s Experiential Learning Theory [62].

Personalizable and Adaptable Components

Previous evidence suggests that CAYP with ADHD would like a mobile app to be personalizable [47]. It is well documented that CAYP with ADHD need to be optimally stimulated to maintain engagement with a task [70-72]. As advised by study participants (5 CAYP, 4 parents, and all clinicians), personalizable avatars that are able to be constantly adapted as and when the users would like could provide them with the stimulation and motivation to remain engaged with the intervention. CAYP with ADHD have also been reported to want to adapt avatars so that they can relate to them [47]. A total of 4 clinicians emphasized the importance of the CAYP being able to relate to the intervention. Support for this can come from a “mini theory” Organismic Integration Theory (OIT), derived from self-determination Theory. OIT emphasizes the importance of relatedness to motivate an individual to behave in a certain way [64].

Psychoeducation Component

Participants (5 CAYP, 5 parents, and all 6 clinicians) wanted CAYP to know more about their ADHD so that they could self-manage it more effectively and so that the CAYP could explain what ADHD means to their peers. This concurs with existing literature where emphasis has been placed on the value of psychoeducation for CAYP with ADHD and their families, as an expert understanding of their condition could lead to more positive individual choices [49,73]. The Health Foundation reports that educating people about their long-term condition can support self-management [74]. Public Heath England [75], along with the Mental Health Taskforce’s 5 Year Forward View for Mental Health [76], states that early intervention avoids CAYP falling into crisis and expensive longer-term interventions into adulthood. This evidence suggests that psychoeducation for CAYP with ADHD as early as possible is vital to help them understand and self-manage their condition. Despite this favorable evidence base for psychoeducation, CAYP with ADHD often do not have access to appropriate psychoeducation, and their understanding of the condition is frequently poor and likely to lower self-esteem.

Integration of Self-Management Strategies

Overall, 5 CAYP, 4 parents, and all 6 clinicians believed the availability of self-management strategies for ADHD could be useful for CAYP with ADHD. Social learning theory states that individuals can learn by imitating others [58]. Animated social scenarios whereby the user can choose alternate endings could enable the user to learn about acceptable behavior in social situations. Bandura’s self-efficacy theory states that “Modeling” can increase self-efficacy [69].

Moreover, 2 clinicians recognized that “social stories” are an effective way to teach CAYP with ASD how to behave appropriately in social situations and are often used in clinical practice [77,78]. Therefore, they believed the proposed animated social scenarios could work well with many CAYP with ADHD, especially those CAYP who have comorbid social skills difficulties.

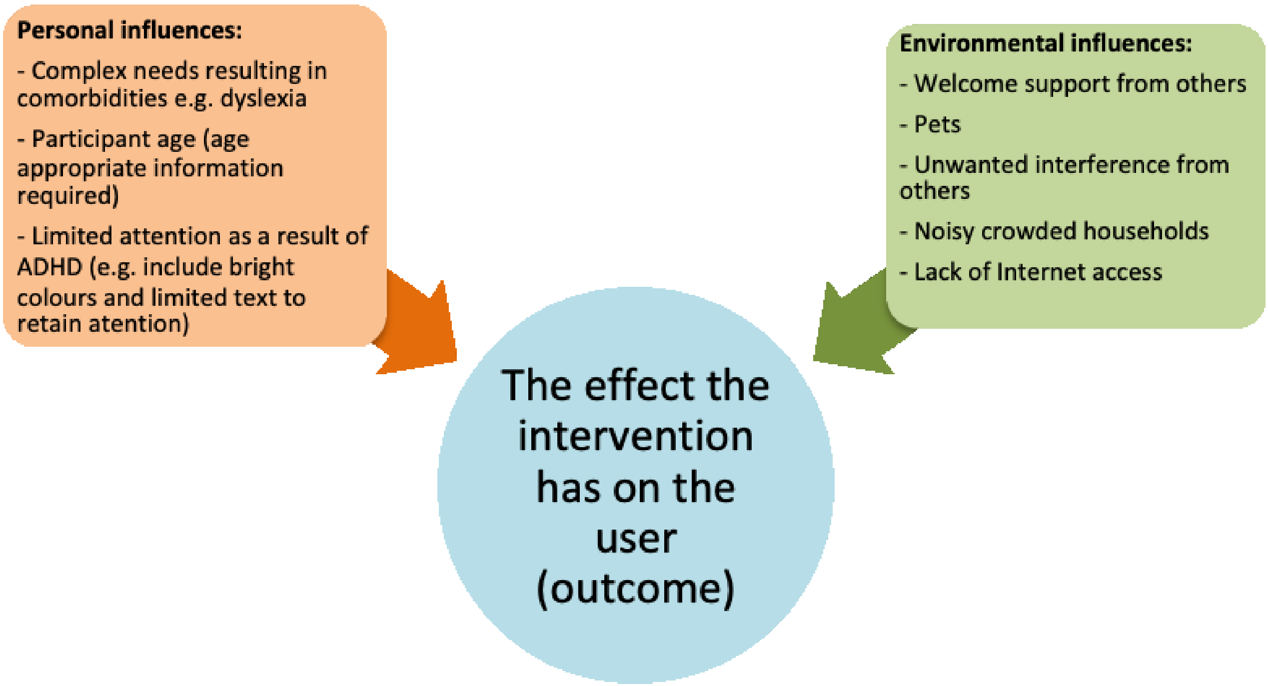
Furthermore, “interpreting physiological signs” is also a stream of Bandura’s theory of self-efficacy [69]. This could have applied to ADHD in CAYP as if the young people can identify when they are likely to feel angry or frustrated, this could be when they apply some self-management strategies to control their behavior, which could lead to an improvement in their self-efficacy.

As CAYP with ADHD can be impulsive, it was requested that interventions should involve a component to help them when they wish to behave impulsively, for example, when they are angry. Support from this may come from the Social Regulation Theory that states CAYP with ADHD lack self-control, which can affect their working memory [59]. This theme is also supported by the CCM, which states that patients should receive support to self-manage their condition [17].

Goal Setting

Overall, 6 parents liked the idea of short-term goal setting within an intervention. Executive dysfunction theory has been applied to ADHD [60], and it states that CAYP with ADHD commonly experience working memory deficits. This is supportive of the fact that goals should be shorter rather than longer-term as the working memory capacities may not enable them to remember requirements to achieve a long-term goal.

Figure 3. Outlines identified environmental and personal contextual factors that could affect the effect (outcome) an intervention has on a user.



Context: Environmental

In accordance with the MRC framework and the International Classification of Function, Disability and Health (ICF), this research found that variable contexts in which an intervention could be delivered could be divided into personal and environmental factors [79]. Figure 3 displays identified environmental and personal contexts as having the potential to change intervention outcomes.

Moreover, 1 example of an environmental context is support from others. A total of 13 participants believed there was value in CAYP with ADHD having support and encouragement (to use an intervention) from their close friends and families. This concurs with the theory of self-efficacy that states that “feedback and persuasion” from significant others, such as family members, can increase one’s self-efficacy [69,80]. Therefore, the support from a close friend or relative when completing such an intervention could help increase the user’s self-efficacy.

Context: Personal

Overall, 1 participant acknowledged that ADHD is a highlight comorbid disorder [81], which includes other conditions such as dyslexia, and 1 provision that could be made is giving users the option to change text background color to aid reading. In addition, optimal stimulation theory states that CAYP with ADHD need to be optimally stimulated to maintain their attention. Therefore, it is important that information and language presented are both age appropriate and interesting to look at, for example, by the use of bright colors.

Strengths, Limitations, and Recommendations

This study has highlighted the importance of considering the variable context in which interventions take place [44]. If research does not consider factors such as the context the intervention is delivered in and the variety in the population, the results could lack reliability and depth [82]. Therefore, this study has provided initial guidelines to assist future technology developers with this process. Furthermore, MRTs were used to underpin the guidelines to help increase their generalizability to more than 1 context. Future research into complex intervention development for any population may wish to adapt the methodology of this study to assist with building an evidence base for the population’s intervention.

Existing evidence is supportive of a psychoeducation component for such interventions [49,73-76]; therefore, future technology should include this component if appropriate.

In addition, the BCW [21] provides a framework for behavior change interventions and the CCM [17] for the care of chronic conditions. During the production of these guidelines, the National Institute for Health and Care Excellence (NICE) released its digital health intervention (DHI) framework [83]. This is an excellent framework that makes a number of detailed recommendations for the development of complex DHIs. These models, and the NICE DHI framework, are valuable for behavior change, chronic care, and complex intervention development, respectively, they are generic models that can be applied to many conditions, not only ADHD. Where the guidelines developed in this instance are partially based upon generic theories such as these, they are also condition specific. This is important for a population with complex needs, such as CAYP with ADHD, as they have needs that cannot be applied to the many conditions the BCW, CCM, and NICE DHI framework target.

Although a sampling frame was adhered to, ensuring a representative sample of this complex population, the qualitative nature of this research meant that the CMOCs for this study were validated and refined using interview data from a small number of participants (n=21). ADHD is a highly complex neurological condition; therefore, 1 intervention will not suit all CAYP with ADHD, or all families and future technological interventions will need to account for this. In addition, for some, these guidelines and subsequent technology development may still not meet their need for personable one-to-one interaction. Furthermore, this study was limited to the views and opinions of CAYP with ADHD, their parents, and specialist clinicians. Game designers and platform developers were not consulted as it was outside the aims and objectives of this study. Future research may benefit from incorporating the views and opinions of these individuals.

Complex interventions for ADHD self-management run the inevitable risk of variable uptake of the intervention among participants [15]; therefore, future attempts should account for this. These guidelines were designed in 2018. Technology is constantly changing and alongside this, so are consumer expectations [43]; therefore, it is important for these guidelines to be reviewed regularly and for future projects to develop complex interventions to be aware of technological developments at the time. Although these guidelines may need reviewing, contexts that complex interventions are delivered in will always be variable; therefore, the methodology adopted for this study could be used beyond the lifetime of the guidelines developed.

Conclusions

This study has adopted the principles of RE [52] to design a set of guidelines that can be used when developing complex, technological interventions that aim to help CAYP aged 8 to 11 years with ADHD self-manage their condition. The guidelines propose helping CAYP aged 8 to 11 years with ADHD understand their condition and providing them with tools to self-manage it more effectively. This concurs with the health foundation’s guide to self-management of long-term conditions [22]. It is anticipated that these guidelines will become a research derived actionable tool [84] in the future for designers to use and maximize the impact they have on the development of technological interventions for this population. It is recommended that a co-design approach should be adopted when designing complex interventions to increase the likelihood of acceptance of the intervention and engagement with the intervention [13,14]. The methodology presented could also be used to stimulate a wide range of stakeholders (service users, clinicians, researchers, and policy makers) to think differently about how interventions for this population, and other populations and age groups, are designed. Beyond the use of these guidelines, future research evaluating the effectiveness of such an intervention must contain large sample sizes and account for the variable contexts interventions are delivered in to ensure that the findings are generalizable. A follow-up period is also essential to evaluate if intervention effects persist over longer periods of time [15,49]. Although these guidelines provide a good theory and evidence basis for the development of a future complex intervention of this nature, it must be acknowledged that it is vital that complex interventions should be codesigned in partnership with key stakeholders to increase the likelihood that the intervention is to be accepted by the intended users [46-48].

Acknowledgments

This research was funded and supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber. The views and opinions expressed are those of the authors and are not necessarily those of the NHS, the NIHR, or the Department of Health, United Kingdom. The authors would like to thank the CAYP, their families, and the clinicians for taking part in this study and sharing their expertise.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The guidelines for the development of technological interventions for Children and young people with attention deficit hyperactivity disorder.

[PDF File (Adobe PDF File), 49KB - jmir_v21i4e12831_app1.pdf]

Multimedia Appendix 2

Refined context mechanism outcome configurations.

[PDF File (Adobe PDF File), 55KB - jmir_v21i4e12831_app2.pdf]

References

1. National Institute for Health and Care Excellence.: NICE; 2018. Attention deficit hyperactivity disorder: diagnosis and management URL: https://www.nice.org.uk/guidance/ng87 [accessed 2019-01-21] [WebCite Cache ID 73vWrOgWg]

2. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int J Epidemiol 2014 Apr;43(2):434-442 [FREE Full text] [doi: 10.1093/ije/dyt261] [Medline: 24464188]
3. National Institute for Health and Care Excellence. Attention deficit hyperactivity disorder: NICE quality standard QS39 URL: https://www.nice.org.uk/guidance/qs39 [accessed 2018-11-14] [WebCite Cache ID 73vX3U2Fy]
4. Erskine HE, Ferrari AJ, Nelson P, Polanczyk GV, Flaxman AD, Vos T, et al. Epidemiological modelling of attention-deficit/hyperactivity disorder and conduct disorder for the Global Burden of Disease Study 2010. J Child Psychol Psychiatry 2013 Dec;54(12):1263-1274. [doi: 10.1111/jcpp.12144] [Medline: 24117530]
5. McGee R, Partridge F, Williams S, Silva PA. A twelve-year follow-up of preschool hyperactive children. J Am Acad Child Adolesc Psychiatry 1991 Mar;30(2):224-232. [doi: 10.1097/00004583-199103000-00010] [Medline: 2016226]
6. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. Mol Psychiatry 2018 Jun 11;17(2):261-248. [doi: 10.1038/s41380-018-0070-0] [Medline: 29892054]
7. Green H, McGinnity Á, Meltzer H, Ford T, Goodman R. NHS Digital. 2004. Mental health of children and young people in Great Britain URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 [accessed 2019-02-19] [WebCite Cache ID 76INxy1u4]
8. Russell G, Rodgers LR, Ukoumunne OC, Ford T. Prevalence of parent-reported ASD and ADHD in the UK: findings from the Millennium Cohort Study. J Autism Dev Disord 2014 Jan;44(1):31-40. [doi: 10.1007/s10803-013-1849-0] [Medline: 23719853]
9. Klein RG, Mannuzza S, Olazagasti MA, Roizen E, Hutchison JA, Lashua EC, et al. Clinical and functional outcome of childhood attention-deficit/hyperactivity disorder 33 years later. Arch Gen Psychiatry 2012 Dec;69(12):1295-1303 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.271] [Medline: 23070149]
10. Parker J, Wales G, Chalhoub N, Harpin V. The long-term outcomes of interventions for the management of attention-deficit hyperactivity disorder in children and adolescents: a systematic review of randomized controlled trials. Psychol Res Behav Manag 2013 Sep 17;6:87-99 [FREE Full text] [doi: 10.2147/PRBM.S49114] [Medline: 24082796]
11. Bussing R, Koro-Ljungberg M, Noguchi K, Mason D, Mayerson G, Garvan CW. Willingness to use ADHD treatments: a mixed methods study of perceptions by adolescents, parents, health professionals and teachers. Soc Sci Med 2012 Jan;74(1):92-100 [FREE Full text] [doi: 10.1016/j.socscimed.2011.10.009] [Medline: 22133584]
12. Beecham J. Annual research review: child and adolescent mental health interventions: a review of progress in economic studies across different disorders. J Child Psychol Psychiatry 2014 Jun;55(6):714-732 [FREE Full text] [doi: 10.1111/jcpp.12216] [Medline: 24580503]
13. Boyd H, McKernon S, Old A. Health Service Co-design: Working with Patients to Improve Healthcare Services : Guide and Toolkit. Auckland, New Zealand: Waitemata District Health Board; 2010.
14. Robert G, Cornwell J, Locock L, Purushotham A, Sturmey G, Gager M. Patients and staff as codesigners of healthcare services. Br Med J 2015 Feb;350:g7714. [doi: 10.1136/bmj.g7714] [Medline: 25670179]
15. 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 2016 Dec 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
16. Eaton S, Roberts S, Turner B. Delivering person centred care in long term conditions. Br Med J 2015 Feb 10;350:h181. [doi: 10.1136/bmj.h181] [Medline: 25670186]
17. Wagner EH, Austin BT, von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [doi: 10.2307/3350391] [Medline: 8941260]
18. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. J Am Medic Assoc 2002 Oct 9;288(14):1775-1779. [Medline: 12365965]
19. Subcommittee on Attention-Deficit/Hyperactivity Disorder, Steering Committee on Quality Improvement and Management, Wolraich M, Brown L, Brown RT, DuPaul G, et al. ADHD: clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. Pediatrics 2011 Nov;128(5):1007-1022. [doi: 10.1542/peds.2011-2654] [Medline: 22003063]
20. Langberg JM, Brinkman WB, Lichtenstein PK, Epstein JN. Interventions to promote the evidence-based care of children with ADHD in primary-care settings. Expert Rev Neurother 2009 Apr;9(4):477-487. [doi: 10.1586/ern.09.5] [Medline: 19344300]
21. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
22. The Health Foundation. 2015. A practical guide to self-management support URL: https://www.health.org.uk/publications/a-practical-guide-to-self-management-support [accessed 2019-01-24] [WebCite Cache ID 75fE9QcYM]
23. Avila-Pesantez D, Rivera L, Vaca-Cardenas L, Aguayo S, Zuñiga L. Towards the improvement of ADHD children through augmented reality serious games: preliminary results. 2018 Presented at: Global Engineering Education Conference (EDUCON); April 17-20, 2018; Santa Cruz de Tenerife, Canary Islands, Spain.

24. Bamford C. Rowan Digital Works.: Rowan University; 2016. The effects of self-monitoring using an iPad to increase on-task behaviors of high school students with Attention Deficit/Hyperactivity Disorder (ADHD) URL: https://rdw.rowan.edu/cgi/viewcontent.cgi?article=2581&context=etd
25. Bul KC, Kato PM, Van der Oord S, Danckaerts M, Vreeke LJ, Willems A, et al. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: a randomized controlled trial. J Med Internet Res 2016 Feb 16;18(2):e26 [FREE Full text] [doi: 10.2196/jmir.5173] [Medline: 26883052]
26. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating health behavior theory and design elements in serious games. JMIR Ment Health 2015;2(2):e11 [FREE Full text] [doi: 10.2196/mental.4133] [Medline: 26543916]
27. Dovis S, van der Oord S, Wiers RW, Prins PJ. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
28. Egeland J, Aarlien AK, Saunes B. Few effects of far transfer of working memory training in ADHD: a randomized controlled trial. PLoS One 2013;8(10):e75660 [FREE Full text] [doi: 10.1371/journal.pone.0075660] [Medline: 24124503]
29. Johnstone SJ, Roodenrys S, Blackman R, Johnston E, Loveday K, Mantz S, et al. Neurocognitive training for children with and without AD/HD. Atten Defic Hyperact Disord 2012 Mar;4(1):11-23. [doi: 10.1007/s12402-011-0069-8] [Medline: 22179720]
30. Klingberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlström K, et al. Computerized training of working memory in children with ADHD--a randomized, controlled trial. J Am Acad Child Adolesc Psychiatry 2005 Feb;44(2):177-186. [doi: 10.1097/00004583-200502000-00010] [Medline: 15689731]
31. Lin C, Yu W, Chen W, Huang C, Lin C. The effect of literacy learning via mobile augmented reality for the students with ADHD and reading disabilities. 2016 Presented at: International Conference on Universal Access in Human-Computer Interaction; July 17-22, 2016; Toronto, Canada.
32. Ota K, DuPaul G. Task engagement and mathematics performance in children with attention-deficit hyperactivity disorderffects of supplemental computer instruction. Sch Psychol Q 2002;17(3):242-257. [doi: 10.1521/scpq.17.3.242.20881]
33. Pinna J. University of Baltimore. 2015. Development and preliminary testing of tablet application to increase reading motivation and summarization for adolescent students with ADHD URL: https://mdsoar.org/bitstream/handle/11603/3790/UB_2015_Pinna_J.pdf?sequence=1&isAllowed=y
34. Rijo R, Costa P, Machado P, Bastos D, Matos P, Silva A, et al. Mysterious Bones Unearthed: development of an online therapeuticserious game for children with attention deficit-hyperactivity disorder. Procedia Comput Sci 2015;64:1208-1216. [doi: 10.1016/j.procs.2015.08.512]
35. Ruiz-Manrique G, Tajima-Pozo K, Montañes-Rada F. Case report: "ADHD Trainer": the mobile application that enhances cognitive skills in ADHD patients.. F1000Res 2014;3:283 [FREE Full text] [doi: 10.12688/f1000research.5689.3] [Medline: 26962432]
36. Shih CH, Wang SH, Wang YT. Assisting children with attention deficit hyperactivity disorder to reduce the hyperactive behavior of arbitrary standing in class with a Nintendo Wii remote controller through an active reminder and preferred reward stimulation. Res Dev Disabil 2014 Sep;35(9):2069-2076. [doi: 10.1016/j.ridd.2014.05.007] [Medline: 24881005]
37. Shrieber B, Seifert T. ResearchGate. 2009. College students with learning disabilities and/or ADHD use of a handheld computer compared to conventional planners URL: https://www.researchgate.net/publication/251295959_College_Students_with_Learning_Disabilities_andor_ADHD_Use_of_a_Handheld_Computer_Compared_to_Conventional_Planners [accessed 2019-02-19] [WebCite Cache ID 76J1DZMiu]
38. Sonne T, Marshall P, Knight F. Using mobile technology interventions to facilitate healthy sleep habits for children with ADHD. Sleep Medicine 2017;40:e181-e182 [FREE Full text] [doi: 10.1016/j.sleep.2017.11.531]
39. van der Oord S, Ponsioen AJ, Geurts HM, Ten BE, Prins PJ. A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: outcome on parent- and teacher-rated executive functioning and ADHD behavior. J Atten Disord 2014 Nov;18(8):699-712. [doi: 10.1177/1087054712453167] [Medline: 22879577]
40. van Dongen-Boomsma M, Vollebregt MA, Buitelaar JK, Slaats-Willemse D. Working memory training in young children with ADHD: a randomized placebo-controlled trial. J Child Psychol Psychiatry 2014 Aug;55(8):886-896. [doi: 10.1111/jcpp.12218] [Medline: 24628438]
41. Weerdmeester J, Cima M, Granic I, Hashemian Y, Gotsis M. A feasibility study on the effectiveness of a full-body videogame intervention for decreasing attention deficit hyperactivity disorder symptoms. Games Health J 2016 Aug;5(4):258-269. [doi: 10.1089/g4h.2015.0103] [Medline: 27304677]
42. Mawson S, Nasr N, Parker J, Davies R, Zheng H, Mountain G. A personalized self-management rehabilitation system with an intelligent shoe for stroke survivors: a realist evaluation. JMIR Rehabil Assist Technol 2016 Jan 7;3(1):e1 [FREE Full text] [doi: 10.2196/rehab.5079] [Medline: 28582250]
43. Parker J, Mawson S, Mountain G, Nasr N, Zheng H. Stroke patients' utilisation of extrinsic feedback from computer-based technology in the home: a multiple case study realistic evaluation. BMC Med Inform Decis Mak 2014 Jun 5;14:46 [FREE Full text] [doi: 10.1186/1472-6947-14-46] [Medline: 24903401]

44. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
45. Dunn DS, Elliott TR. The place and promise of theory in rehabilitation psychology researc. Rehabil Psychol 2008 Aug;53(3):254-267 [FREE Full text] [doi: 10.1037/a0012962] [Medline: 19649146]
46. Powell L, Joddrell P, Parker J. Involving users in the evaluation of apps for specific health conditions. Stud Health Technol Inform 2017;242:646-653. [doi: 10.3233/978-1-61499-798-6-646] [Medline: 28873866]
47. Powell L, Parker J, Robertson N, Harpin V. Attention deficit hyperactivity disorder: is there an app for that? suitability assessment of apps for children and young people with ADHD. JMIR Mhealth Uhealth 2017 Oct 4;5(10):e145 [FREE Full text] [doi: 10.2196/mhealth.7371] [Medline: 28978497]
48. Powell L, Parker J, Harpin V. ADHD: is there an app for that? A suitability assessment of apps for parents of children and young people with ADHD. JMIR Mhealth Uhealth 2017 Oct 13;5(10):e149 [FREE Full text] [doi: 10.2196/mhealth.7941] [Medline: 29030325]
49. Powell L, Parker J, Harpin V. What is the level of evidence for the use of currently available technologies in facilitating the self-management of difficulties associated with ADHD in children and young people? A systematic review. Eur Child Adolesc Psychiatry 2018 Nov;27(11):1391-1412. [doi: 10.1007/s00787-017-1092-x] [Medline: 29222634]
50. Keller C, Gäre K, Edenius M, Lindblad S. Designing for complex innovations in health care: design theory and realist evaluation combined. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. 2009 Presented at: DESRIST'09; May 7-8, 2009; Philadelphia, Pennsylvania.
51. Pearson M, Brand SL, Quinn C, Shaw J, Maguire M, Michie S, et al. Using realist review to inform intervention development: methodological illustration and conceptual platform for collaborative care in offender mental health. Implement Sci 2015 Sep 28;10:134 [FREE Full text] [doi: 10.1186/s13012-015-0321-2] [Medline: 26415961]
52. Pawson R, Tilley N. Realistic Evaluation. London: Sage Publications; 1997.
53. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC medicine 2016;14(1):96. [doi: 10.1186/s12916-016-0643-1] [Medline: 29072890]
54. Jagosh J, Bush PL, Salsberg J, Macaulay AC, Greenhalgh T, Wong G, et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. BMC Public Health 2015 Jul 30;15:725 [FREE Full text] [doi: 10.1186/s12889-015-1949-1] [Medline: 26223523]
55. Pavlov IP. The Work of the Digestive Glands: Lectures by Professor J P Pawlow. London: C Griffin; 1910.
56. Skinner B. About Behaviourism. New York: MacMillan; 1974.
57. Leuba C. Toward some integration of learning theories: the concept of optimal stimulation. Psychol Rep 1955;1(1):27-33. [doi: 10.2466/pr0.1955.1.g.27]
58. Bandura A. Social Foundations of Thought and Action: A Social-Cognitive View. Englewood Cliffs, NJ: Prentice Hall; 1986.
59. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [doi: 10.1037/0033-2909.121.1.65] [Medline: 9000892]
60. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biol Psychiatry 2005 Jun 1;57(11):1336-1346. [doi: 10.1016/j.biopsych.2005.02.006] [Medline: 15950006]
61. Sagvolden T, Johansen E, Aase H, Russell V. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. Behav Brain Sci 2005 Jun;28(3):397-419. [doi: 10.1017/S0140525X05000075] [Medline: 16209748]
62. Dewey J. Experience and Education. New York: Simon and Schuster; 1997.
63. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066X.44.9.1175] [Medline: 2782727]
64. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68] [Medline: 11392867]
65. OpenDataCommunities. 2015. Indices of Deprivation 2015 URL: http://dclgapps.communities.gov.uk/imd/idmap.html [accessed 2018-11-14] [WebCite Cache ID 73vXGIQc1]
66. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
67. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. Br Med J 2000 Jan 1;320(7226):50-52 [FREE Full text] [doi: 10.1136/bmj.320.7226.50] [Medline: 10617534]
68. Yin R. Case Study Research: Design and Methods. Beverly Hills, CA: Sage Publications; 2003.
69. Bandura A. Cultivate self-efficacy for personal and organizational effectiveness. In: Locke EA, editor. Handbook of Principles of Organizational Behavior: Indispensable Knowledge for Evidence ‐ Based Management, Second Edition. West Sussex, UK: John Wiley & Sons Ltd; 2000:179-200.

70. Banaschewski T, Brandeis D, Heinrich H, Albrecht B, Brunner E, Rothenberger A. Association of ADHD and conduct disorder--brain electrical evidence for the existence of a distinct subtype. J Child Psychol Psychiatry 2003 Mar;44(3):356-376. [doi: 10.1111/1469-7610.00127] [Medline: 12635966]
71. Brandeis D, Banaschewski T, Baving L, Georgiewa P, Blanz B, Warnke A, et al. Multicenter P300 brain mapping of impaired attention to cues in hyperkinetic children. J Am Acad Child Adolesc Psychiatry 2002;41(8):990-998. [doi: 10.1097/00004583-200208000-00018] [Medline: 12162635]
72. Zentall S. Theory and evidence based strategies for children with attentional problems. Psychol Sch 2005;42(8):821-836 [FREE Full text] [doi: 10.1002/pits.20114]
73. Ferrin M, Taylor E. Child and caregiver issues in the treatment of attention deficit–hyperactivity disorder: education, adherence and treatment choice. Future Neurol 2011 May;6(3):399-413 [FREE Full text] [doi: 10.2217/fnl.11.5]
74. The Health Foundation. 2011. Helping people help themselves URL: https://www.health.org.uk/sites/health/files/HelpingPeopleHelpThemselves.pdf [accessed 2018-11-14] [WebCite Cache ID 73vXMdmnP]
75. GOV.UK. England: Public Health England; 2016. The mental health of children and young people in England URL: https://www.gov.uk/government/publications/improving-the-mental-health-of-children-and-young-people [WebCite Cache ID 73vXWqre7]
76. Mental Health Taskforce. NHS England. 2016. The five year forward view for mental health URL: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf [WebCite Cache ID 75fEEDYvE]
77. Acar C, Tekin-Iftar E, Yikmis A. Effects of mother-delivered social stories and video modeling in teaching social skills to children with autism Spectrum disorders. J Spec Educ 2017;50(4):215-226 [FREE Full text] [doi: 10.1177/0022466916649164]
78. Smith C. Writing and Developing Social Stories. London: Routledge; 2017.
79. World Health Organisation. 2001. International classification of functioning, disabilityhealth: ICF URL: https://www.who.int/classifications/icf/en/ [WebCite Cache ID 73vXatKe3]
80. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
81. Biederman J, Newcorn J, Sprich S. Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. Am J Psychiatry 1991 May;148(5):564-577. [doi: 10.1176/ajp.148.5.564] [Medline: 2018156]
82. Rycroft-Malone J, Fontenla M, Bick D, Seers K. A realistic evaluation: the case of protocol-based care. Implement Sci 2010 May 26;5:38 [FREE Full text] [doi: 10.1186/1748-5908-5-38] [Medline: 20504293]
83. National Institute for Health and Care Excellence. 2018. Evidence Standards Framework for Digital Health Technologies URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [WebCite Cache ID 75fELArTW]
84. Hampshaw S, Cooke J, Mott L. What is a research derived actionable tool, and what factors should be considered in their development? A Delphi study. BMC Health Serv Res 2018 Sep 27;18(1):740 [FREE Full text] [doi: 10.1186/s12913-018-3551-6] [Medline: 30261925]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ASD: autism spectrum disorder
BCW: Behavior Change Wheel
CAYP: children and young people
CCM: Chronic Care Model
CMOC: context mechanism outcome configuration
COM-B: Capability Opportunity Motivation-Behavior
DHI: digital health intervention
MRC: Medical Research Council
MRTs: middle-range theories
NHS: National Health Service
NIHR: National Institute for Health Research
NICE: National Institute for Health and Care Excellence
OIT: Organismic Integration Theory
RCT: randomized controlled trial
RE: realist evaluation

Edited by G Eysenbach; submitted 16.11.18; peer-reviewed by B Chaudhry, J Appalasamy, Y Qiu, K Bul; comments to author 17.01.19; revised version received 24.01.19; accepted 28.01.19; published 03.04.19.

Please cite as:
Powell L, Parker J, Harpin V, Mawson S
Guideline Development for Technological Interventions for Children and Young People to Self-Manage Attention Deficit Hyperactivity Disorder: Realist Evaluation
J Med Internet Res 2019;21(4):e12831
URL: https://www.jmir.org/2019/4/e12831/
doi:10.2196/12831
PMID:30942692

Original Paper

Implementation of a Novel Electronic Patient-Directed Smoking Cessation Platform for Cancer Patients: Interrupted Time Series Analysis

Meredith Elana Giuliani[1], MBBS, MEd, FRCPC; Geoffrey Liu[1], MD, MS; Wei Xu[1], PhD; Mihaela Dirlea[1], MSW; Peter Selby[2], MBBS, MHSc; Janet Papadakos[1], PhD, MEd; Nazek Abdelmutti[1], MSc; Dongyang Yang[1], MMath; Lawson Eng[3], MD; David Paul Goldstein[1], MD, MSc; Jennifer Michelle Jones[1], PhD

[1]Princess Margaret Cancer Centre, Toronto, ON, Canada

[2]Centre for Addiction and Mental Health, Toronto, ON, Canada

[3]Division of Medical Oncology and Hematology, Department of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jennifer Michelle Jones, PhD
Princess Margaret Cancer Centre
610 University Ave
Toronto, ON, M5G 2C1
Canada
Phone: 1 4167864583
Email: jonesbernard5@gmail.com

Abstract

Background: Continued smoking in cancer patients undergoing treatment results in significantly higher rates of treatment toxicities and persistent effects, increased risk of recurrence and second malignancy, and increased all-cause mortality. Despite this, routine tobacco use screening and the provision of smoking cessation treatment has yet to be implemented widely in the cancer setting.

Objective: The objective of this study was to implement and evaluate the adoption and impact of an innovative Smoking Cessation e-referral System (CEASE) to promote referrals to smoking cessation programs in cancer patients.

Methods: A patient-directed electronic smoking cessation platform (CEASE) was developed to promote smoking screening and referral and implemented at 1 of Canada's largest cancer centers. The implementation and evaluation were guided by the Ottawa Model of Research Use. An interrupted time series design was used to examine the impact of CEASE on screening rates, referrals offered, and referrals accepted compared with a previous paper-based screening program. A subsample of smokers or recent quitters was also assessed and compared pre- and postimplementation to examine the effect of CEASE on subsequent contact with smoking cessation programs and quit attempts.

Results: A total of 17,842 new patients attended clinics over the 20-month study period. The CEASE platform was successfully implemented across all disease sites. Screening rates increased from 44.28% (2366/5343) using the paper-based approach to 65.72% (3538/5383) using CEASE (*P*<.01), and referrals offered to smokers who indicated interest in quitting increased from 18.6% (58/311) to 98.8% (421/426; *P*<.01). Accepted referrals decreased from 41% (24/58) to 20.4% (86/421), though the overall proportion of referrals generated from total current/recent tobacco users willing to quit increased from 5.8% (24/414) to 20.2% (86/426) due to the increase in referrals offered. At 1-month postscreening, there was no significant difference in the proportion that was currently using tobacco and had not changed use in the past 4 weeks (pre: 28.9% [24/83] and post: 28.8% [83/288]). However, contact with the referral program increased from 0% to 78% in the postCEASE cohort (*P*<.001).

Conclusions: CEASE is an innovative tool to improve smoking screening and can be implemented in both a time- and cost-effective manner which promotes sustainability. CEASE was successfully implemented across all clinics and resulted in improvements in overall screening and referral rates and engagement with referral services.

(J Med Internet Res 2019;21(4):e11735) doi:10.2196/11735

KEYWORDS

neoplasms; smoking cessation; implementation science; quality improvement

Introduction

Background

It is widely recognized that smoking cessation will decrease the risk of developing certain cancers [1]. In addition, there is a large body of evidence which demonstrates that continued smoking after a cancer diagnosis leads to significantly higher rates of treatment toxicities and persistent effects [1-7], increased the risk of recurrence and second malignancy [1,3,4,8-11], and increased all-cause mortality [1-3,8-10,12-14]. Unfortunately, 20% to 25% of individuals will continue to smoke after a cancer diagnosis and throughout treatment [15,16].

Screening and advice to patients on smoking cessation can be an effective first step to smoking cessation and increases the likelihood that a patient will attempt to quit and be successful [17]. Over the past 5 years, a number of leading cancer organizations have released policy statements which recommend timely and cost-effective assessment of tobacco use and the provision of smoking cessation assistance for all cancer patients [2,17,18]. In 2016, the National Comprehensive Cancer Network released its Clinical Practice Guidelines in Oncology for Smoking Cessation, which highlighted the importance of smoking cessation and recommended that all cancer patients be evaluated and assessed for smoking status, offered counseling on quitting, and provided with tailored cessation services [18]. In Canada, the CAN-ADAPTT (Canadian Action Network for the Advancement, Dissemination and Adoption of Practice-informed Tobacco Treatment) Clinical Practice Guideline for Smoking Cessation [19] recommends that (1) tobacco use status should be updated for all patients by their health care providers (HCPs) on a regular basis and that HCPs should clearly advise to quit and assess the willingness to begin treatment to achieve abstinence (ask, advise, assess), (2) every tobacco user who expresses willingness to begin treatment to quit be offered assistance (assist), and (3) HCPs conduct regular follow-up to assess response and are encouraged to refer patients to relevant resources as part of the provision of treatment (arrange). Cancer Care Ontario (CCO), an agency which oversees the quality of cancer services in the province of Ontario, conducted an environmental scan to determine the need for a standardized approach of smoking cessation programs within the cancer system and subsequently developed a Smoking Cessation Advisory Committee, which established a plan and provided guidance for the implementation of a smoking cessation program in regional cancer centers within Ontario. In the 2011-2015 Cancer Plan, CCO mandated that all new cancer patients be screened for smoking status and that recent or current smokers be advised to quit and assisted with quitting [20].

Despite the myriad of benefits of smoking cessation to cancer patients [3,11,21-24] and the fact that many newly diagnosed cancer patients are motivated to quit smoking and are open to discussions on how to do this [25-27], there remain significant challenges in terms of the implementation of these recommendations and strategies in oncology care settings [20,28]. Consequently, the majority of cancer patients are not screened for smoking status and/or referred to cessation services [2]. This knowledge-to-practice gap in screening and referrals is likely multifactorial [29] and requires innovative and sustainable approaches that consider the realities of the clinical environment and can efficiently screen and refer large volumes of cancer patients.

Objective and Specific Aims

The objective of this study was to implement and evaluate the adoption and impact of an innovative Smoking *C* essation *e*-referra *l* *S* yste m (CEASE) to promote referrals to smoking cessation programs in cancer patients. The specific aims of the study were (1) to facilitate the adoption of CEASE in promoting smoking screening and referral to cessation programs and (2) to evaluate the impact of CEASE on screening and referral patterns. We hypothesized that the CEASE system would be successfully implemented and acceptable to patients, would result in increased screening and referral rates, and would subsequently result in increased interactions with smoking cessation programs.

Methods

Our project followed the Standards for Quality Improvement Reporting Excellence 2.0 guidelines for study design and analysis [30] and was guided by the Ottawa Model of Research Use (OMRU) [31]. This study was reviewed and approved by the University Health Network Research Ethics Board (#15-8974 CE).

Context

The Princess Margaret Cancer Centre (PM) is 1 of the 14 regional cancer centers in Ontario and the largest single-site cancer hospital in Canada. It comprises 12 cancer site groups, 26 specialty clinics, and approximately 3000 staff who see over 400,000 patient visits each year. In 2017, there were approximately 18,000 new patients registered at PM.

In an effort to align with CCO recommendations, a paper-based screening program was implemented throughout PM between 2014 and 2016 with the goal that every new patient at PM be screened for smoking status and provided with a smoking cessation referral or resource when appropriate. At the time of clinic registration, newly diagnosed cancer patients were identified by the patient flow coordinator within each clinic and provided with a paper screening form to complete and return. The form queried patients on their smoking habits and their interest in smoking cessation. Following this, the screening form was placed in the chart to be reviewed by an HCP during the appointment. If the patient was interested in cessation programs, a referral sheet for the Nicotine Dependence Clinic (Centre for Addiction and Mental Health), Smoker’s Helpline, or the hospital pharmacy was completed by the HCP. If the patient did not want a referral, an information pamphlet and referral numbers for the Smoker’s Helpline and pharmacy were provided. The program’s performance was monitored by collecting the number of eligible patients screened and offered a referral each month. This program resulted in an average of 55% (range across clinics 10%-90%) of all new patients screened for smoking status, with 60% to 70% indicating interest in quitting at the time of screening, but only 20% of patients who indicated interest in quitting received a referral by their HCP to a smoking cessation program.

Our initial research [32-39] and subsequent implementation activities allowed for the identification of enablers for tobacco screening and referral, which included the CCO framework and mandate, PM leadership support, support from the clinical teams, as well as high motivation from the patients in terms of interest in smoking cessation (potential adopters). However, time constraints in already overloaded oncology clinics as well as a lack of familiarity with cessation resources remained significant barriers that resulted in substandard screening and referral rates even when a patient indicated interest in quitting (practice environment; OMRU stage 1). On the basis of these findings, we have adapted our approach and developed an electronic patient-driven model to enable systematic screening and patient self-referral (CEASE). The CEASE model provided a solution to address the time constraints of overloaded oncology clinics as well as a lack of familiarity with smoking cessation resources by oncology HCPs. Furthermore, engagement of patients in their own health care may have considerable potential to achieve beneficial outcomes and can be an important and effective strategy to target knowledge-to-care gaps [40,41].

Intervention

CEASE is delivered electronically to newly diagnosed cancer patients on a tablet at the point of care and consists of 3 elements that align with the CAN-ADAPTT Clinical Practice Guideline for Smoking Cessation: (1) a patient-reported smoking assessment tool (ask, assess); (2) brief, standardized patient education regarding smoking (advise), and (3) a simple patient-directed automatic referral system (assist, arrange). On the basis of the screening status (smoker or ecently quit [<6 months] or nonsmoker), a tailored response is generated (see Figure 1). Referrals are automatically sent through the tablet, and the referral sources call within a week to follow-up with the patient. Data from CEASE are archived within the electronic patient record.

Implementation

To facilitate implementation, we employed multiple enabling and reinforcing strategies based on the Awareness-to-Adherence Model of behavior change [42,43]. The preliminary consultation, diffusion, and dissemination strategy were conducted between July 2015 and October 2015 and included the following strategies to promote *awareness* and *agreement*: (1) gathering feedback from stakeholders (including health care team and patients) on workflow to fine-tune the implementation approach; (2) interviews with patients to gather feedback on the CEASE platform and interface, patient-directed messages, and patient education materials; and (3) presentations to the site teams at weekly tumor boards and rounds (and copy via email) to increase awareness, target attitudes, and to provide an introduction to the CEASE platform and workflow [44]. Following final revisions to the implementation approach, CEASE was rolled out in a step-wise process from October 2015 to January 2016 (initial implementation). Following initial implementation, we employed the following strategies to facilitate *adoption* and *adherence*:

1. Audit and feedback were conducted in each clinic to document if CEASE had been completed in the target population. The performance metrics were then provided to disease site teams (and compared with others) for discussion and to develop solutions to any barriers.
2. Reminders regarding the CEASE program were integrated into routine clinical care team meetings or rounds and sent via email to HCPs as part of the stimulus to the change in practice expected.
3. Information posters were developed through the Cancer Education Program to inform patients about CEASE and encourage them to complete the tablet-based tool. Final workflow and system changes were completed in May 2016.

Outcomes

Process-of-Care Outcomes

An interrupted time series design was used to examine the implementation and impact of CEASE on screening rates, referrals offered, and referrals accepted. The study included 20 monthly intervals: 6 months before implementation (April to September 2015; pre), 8 months during a transition period to accommodate a gradual implementation across all tumor sites (October 2015 to May 2016), and 6 months after full implementation (June 2016 to November 2016; post).

Patient-Reported Outcomes

To evaluate the effect of CEASE on subsequent contact with smoking cessation programs and subsequent quit attempts, 1 month following the initial screening (either prepaper screening or postCEASE), a subsample of patients who indicated they were current or recent smokers were sent a follow-up questionnaire to assess uptake of referrals, quit attempts, and reassess smoking status.

Analysis

Segmented regression was used to assess the impact of the changes on 4 prespecified process-of-care outcomes [44,45]: (1) the proportion screened among all the new patients, (2) the proportion offered referral among total of current smokers and the ones who quit smoking in less than 6 months, (3) the proportion of referral accepted among all the offered referral patients, and (4) the proportion of patients willing to quit among the total of current smokers and the ones who quit smoking in less than 6 months. The segmented regression analysis estimates the interaction terms between the implementation of CEASE and time. As there are 3 time segments, we termed the prepaper screening stage the *pre* (ie, preintervention) period; the second time segment was during the roll-out implementation of CEASE; and the third time segment was after CEASE had been fully implemented across the cancer center, termed *post* for the postintervention period. In the models, the binary regression term *intervention 1* represented the comparison of screening after the start of CEASE (October 2015) versus the *pre*period; the binary regression term *intervention 2* represents the comparison of the *post*period with the time segments before the *post*period.

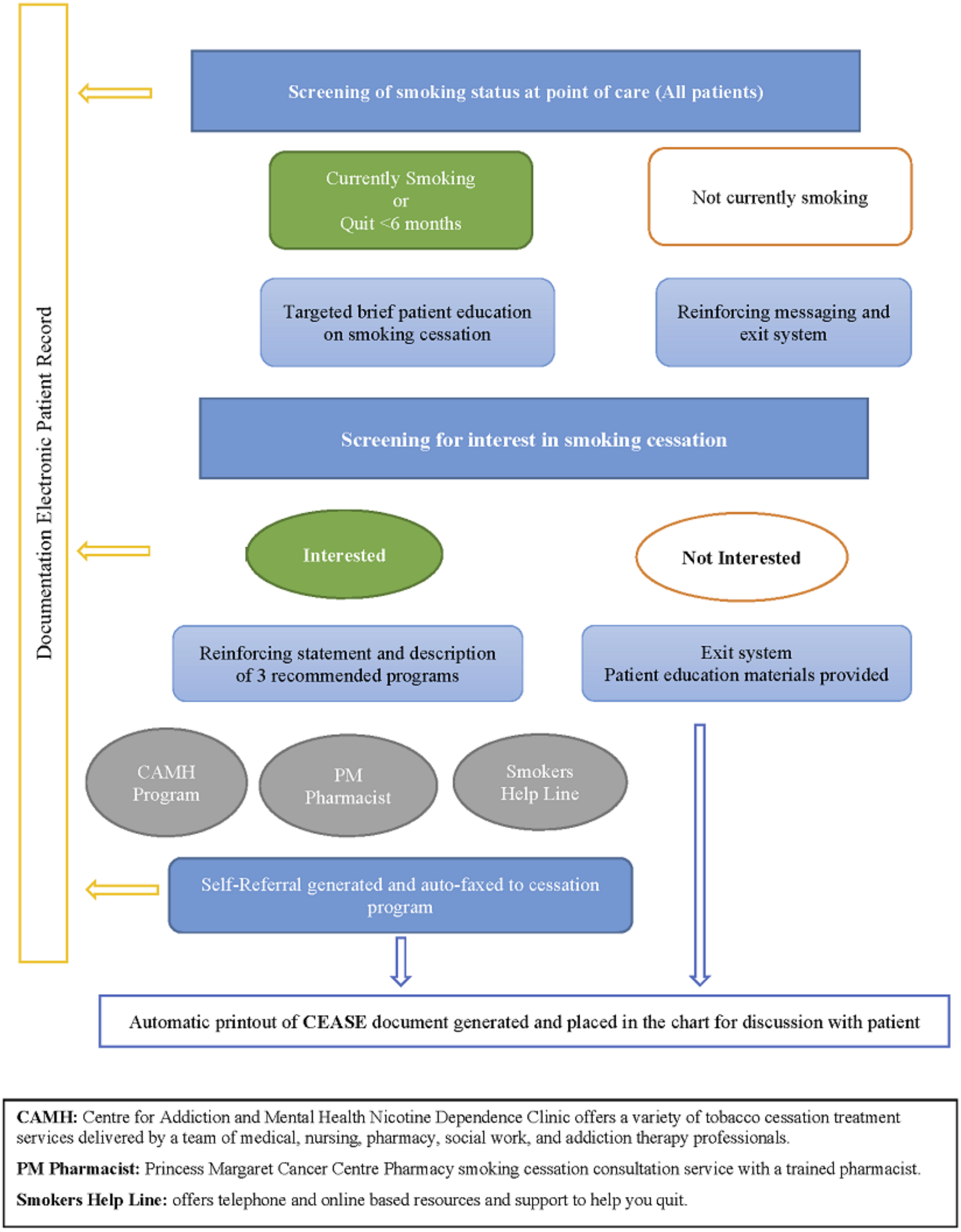
For each outcome, the segmented regression model had the following form:

$$Outcome_t=\beta_0+\beta_1 x\ \ Time+\beta_2 X\ \ Intervention\ \ 1_t+\beta_3 X\ \ Time\ \ after\ \ intervention\ \ 1_t+\beta_4 X\ \ Intervention\ \ 2_t+\ Intervention\ 2_t+\beta_5 X\ Time\ after\ intervention\ 2_t+\ {}_t$$

where $_t$ is the error term following an auto-regression model adjusting for serial correlation [44,45], and β_0 is the intercept for prepaper screening stage, $\beta_0+\beta_2$ is the intercept for *intervention 1* stage, and $\beta_0+\beta_2+\beta_4$ is the intercept for *intervention 2* stage, whereas β_1 is the slope for prepaper screening stage, $\beta_1+\beta_3$ is the slope for *intervention 1* stage, and $\beta_1+\beta_3+\beta_5$ is the slope for *intervention 2* stage. The segmented regression was conducted using PROC AUTOREG in SAS version 9.4 [46].

Prepost, self-reported patient outcome data were compared using Chi-square tests [47] using SPSS version 24.0. In addition, 2-sided tests were conducted, and the statistical significance was set at $P<.05$.

Figure 1. Smoking cessation screening and referral process flow.



CAMH: Centre for Addiction and Mental Health Nicotine Dependence Clinic offers a variety of tobacco cessation treatment services delivered by a team of medical, nursing, pharmacy, social work, and addiction therapy professionals.

PM Pharmacist: Princess Margaret Cancer Centre Pharmacy smoking cessation consultation service with a trained pharmacist.

Smokers Help Line: offers telephone and online based resources and support to help you quit.

Results

A total of 17,842 new patients attended clinics over the study period, including 5343 during the 6-month preintervention period, 7116 during the 8-month implementation period, and 5383 during the 6-month postimplementation period. Figure 2 presents the screening methods used (paper vs CEASE) over the course of the study. By the end of the implementation period and throughout the postimplementation period, only 36 out of the 3538 (1.01%) patients who were screened received the screening on paper.

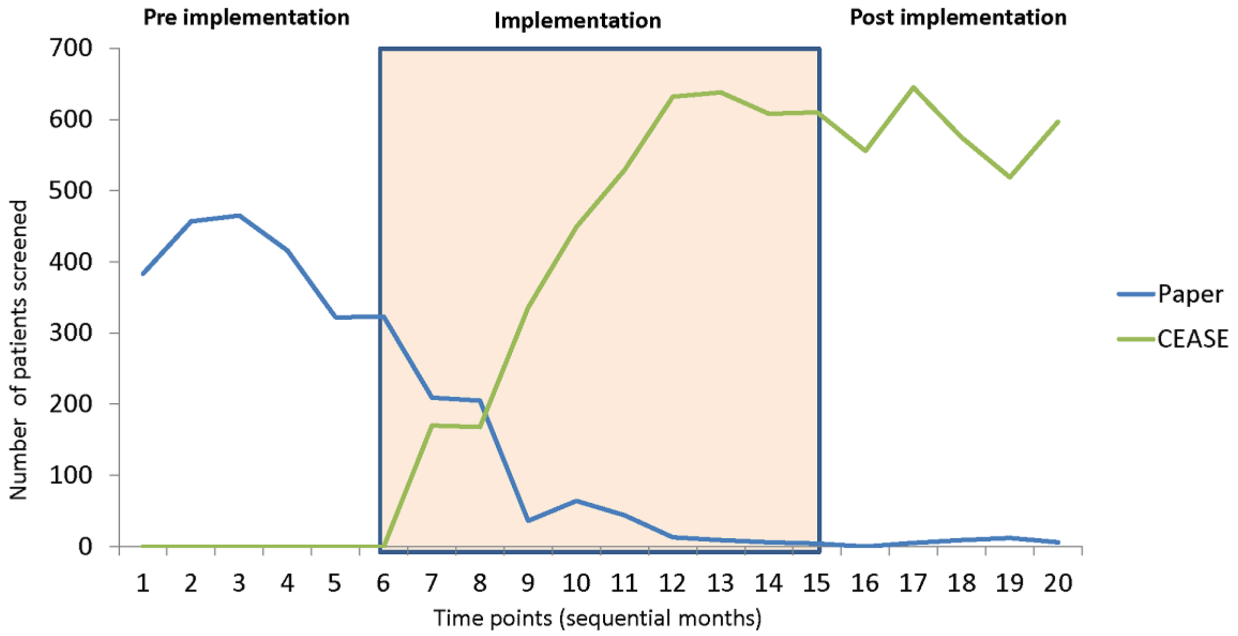
Process of Care Outcomes

Multimedia Appendices 1 and 2 present the time series for each outcome indicator, with fitted trends. The final model fit was found to be adequate. The estimated coefficients from the segmented regression analyses are presented in Multimedia Appendix 3. Screening rates increased from preimplementation at 44.28% (2366/5343) using the paper-based approach to 65.72% (3538/5383) at postimplementation using CEASE (*P*<.01). Referrals offered to smokers who indicated interest in quitting increased from 18.6% (58/311) to 98.8% (421/426; *P*<.01). Accepted referrals decreased from 41% (24/58) to 20.4% (86/421), though the overall proportion of referrals generated from total current or recent tobacco users willing to quit increased from 7.7% (24/414) to 20.2% (86/426) due to the increase in referrals offered.

Pre-Post Self-Report Patient-Reported Data

A total of 29.7% (83/279) of surveys were completed and returned during the preimplementation (paper) phase and 41.9% (288/686) during the postimplementation (CEASE) period. The 2 samples did not differ on any demographic variables. At 1-month postscreening, 24 of 83 (29%) patients in the precohort were still smoking and 83 of 288 (28.8%) patients in the postcohort reported that they were currently using tobacco; 20 of 88 (23%) patients in the precohort and 80 of 288 (27.7%) patients in the postcohort were ok currently using tobacco but had reduced tobacco use over the past 4 weeks; and 37 of 88 (42%) patients in the precohort and 101 of 288 (35.1%) patients in the postcohort reported that they had stopped smoking (pre: 42% and post: 35%). A total of 47% (41/88) of precohort and 76.4% (220/288) of postcohort respondents remembered completing a screening questionnaire about tobacco use at their first visit at PM (*P*<.001). In the precohort, 24% reported receiving a referral from their HCP but none (0%) reported that they had been contacted or had followed up with the referral program. In the postcohort phase, 24% reported that they had accepted a referral through the CEASE program, and of these, 78% had been contacted or followed up with the referral program (*P*<.001).

Figure 2. Screening method (paper and CEASE) over time. CEASE: Cessation e-referral System.



Discussion

Principal Findings

The implementation of CEASE was shown to be feasible and sustainable within a large cancer center with a high volume of patients and supports the use of a technology- and patient-mediated implementation approach. CEASE was also successfully integrated into the electronic medical record in a sustainable fashion. Following implementation, only 1% of patients refused to use the CEASE platform and asked for the paper version, a finding that supports its acceptability. Compared with paper-based screening and HCP-dependent referral, the implementation of the CEASE platform significantly increased screening rates from 44% to 66% and referral rates from 19% to 99%.

It is clear from the recommendations of multiple leading cancer organizations that smoking cessation support is a critical component of a quality cancer program [2,17,18]. Despite this, cancer patients are unlikely to have their smoking systematically assessed and managed [48]. The CEASE platform is able to address all the recommended standards for assessing the smoking status in cancer patients and ensuring that the recommendations for access to cessation services are followed [2]. The platform capitalizes on the use of technology to overcome the most common and problematic barriers to cessation screening [20,28,49] and allows for universal screening and an automatic referral to existing resources and specialized providers. In addition, the use of iPads to deliver the intervention at the point of care, integrated with the hospital electronic medical records, and use of customized logic to personalize responses makes it a highly personalized, efficient, scalable, and sustainable initiative.

We found fairly low interest in participating in formal smoking cessation programs, which has been documented elsewhere and may be due to a number of factors such as low motivation and/or confidence and stigma associated with seeking smoking cessation treatment [46]. Many patients want to quit on their own and, therefore, feel that formal smoking cessation treatment is not needed [50], despite strong evidence that treatment for tobacco dependence is associated with significantly higher long-term quit rates [51]. Interestingly, although screening and referral rates improved significantly with the CEASE program, the proportion of patients who accepted referrals decreased from 41%, when the HCP offered the referral, to 20% when offered through CEASE. Although the reason for this finding is not entirely clear, it is possible that patients are more motivated to quit when they receive this recommendation from their treating oncologist. Advice from an HCP can be a very powerful motivator for behavior change [52]. On the other hand, patients may feel pressured to accept the referral because they worry that it may impact care if they say *no* [53]. Although social desirability is one possible reason for the drop in acceptance of the referral, many patients who refused referrals on the CEASE platform indicated that they would like to quit on their own (they had the option to indicate this). Although the proportion of patients who would like to try quitting on their own likely did not change pre or postimplementation, it may be that HCPs would have discouraged this and tried to provide more support through a referral. Encouragingly, our preliminary postsurvey data found similar overall quit rates between the paper-based and CEASE programs, which suggests that CEASE does not impact quit rates negatively (or positively) despite the higher rate of referral acceptance in the paper-based model. However, the long-term impact of CEASE on cessation rates could not be determined from this study and future work with longer follow-up at 12 months or longer is needed. From a population health perspective, it is important to also note that, despite the finding that less patients accepted a referral through the CEASE system, the overall (absolute) proportion of patients who received a referral actually increased from 8% to 20% due to the fact that all current or recent tobacco users were offered a referral on the CEASE system compared with only 19% of those with the paper-based HCP-dependent system. It is also important to consider that none of the precohort patients who reported receiving a referral through their HCP remembered being contacted or following up with the smoking cessation program, compared with 78% of those who generated their own referral through the CEASE program. A manual system that relies on an HCP to generate and send referrals may seem fairly straightforward, but it requires a clear protocol and workflow, which can be difficult in very busy oncology clinics.

Strengths and Limitations

The results of this study need to be considered within the context of its limitations. To begin, although we are unaware of any threats to validity, it is possible that events or initiatives outside of the control of the research team occurred at the same time as the intervention, though we are not aware of any other changes to the clinic set-up during the study period or efforts to address smoking screening or cessation. There are also limitations of the platform itself. First, the CEASE intervention was created and implemented in English only. Therefore, the generalizability of these results to patients whose primary language is not English is not known. The electronic interface is amenable to translation, and this would be a valuable future contribution to the literature. In addition, this study was not designed to assess the long-term impact of CEASE on cessation rates, and the follow-up survey was only administered 1 month after screening, which may not allow enough time to properly assess quit rates and attempts. However, it is feasible within the CEASE platform to program reassessments at prespecified time intervals. In this regard, the team recently received approval and implemented routine reassessments of patients who indicated that they were current or recent tobacco users at diagnosis.

Despite these limitations, our study has used an established methodology and knowledge translation framework over a substantial period with a large patient population. To our knowledge, this is the first such study in cancer patients, and the data provided here may guide the development and implementation of smoking cessation screening programs in other cancer and possibly noncancer programs. For example, the CEASE program is now being expanded to the Ontario lung cancer screening program and is being piloted in 2 other cancer programs in Toronto. There has also been interest in noncancer programs (ie, cardiovascular).

Conclusions

In conclusion, a large majority of newly diagnosed cancer patients are interested in quitting smoking. Electronic, patient-driven screening and referrals via CEASE were successfully implemented across all clinics and resulted in improvements in overall screening and referral rates and engagement with referral services. This represents a sustainable strategy for routine cessation services in cancer care.

Acknowledgments

This study was funded by an Ontario Institute for Cancer/Cancer Care Ontario KT research grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Percentage of patients screened and offered a referral.

[PNG File, 95KB - jmir_v21i4e11735_app1.png]

Multimedia Appendix 2

Percentage of patients who accepted the referral (from those offered a referral) and percentage of referrals generated (from all smokers or current quitters).

[PNG File, 101KB - jmir_v21i4e11735_app2.png]

Multimedia Appendix 3

Estimated coefficients from segmented regression analyses.

[DOCX File, 32KB - jmir_v21i4e11735_app3.docx]

References

1. US Department of Health and Human Services. The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA; 2014. URL: https://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf
2. Toll BA, Brandon TH, Gritz ER, Warren GW, Herbst RS, AACR Subcommittee on TobaccoCancer. Assessing tobacco use by cancer patients and facilitating cessation: an American Association for Cancer Research policy statement. Clin Cancer Res 2013 Apr 15;19(8):1941-1948 [FREE Full text] [doi: 10.1158/1078-0432.CCR-13-0666] [Medline: 23570694]
3. Florou AN, Gkiozos IC, Tsagouli SK, Syrigos KN. Clinical significance of smoking cessation in patients with cancer: a 30-year review. Respir Care 2014;59(12):36-30. [doi: 10.4187/respcare.02559] [Medline: 25185148]
4. Warner DO. Perioperative abstinence from cigarettes: physiologic and clinical consequences. Anesthesiology 2006 Feb;104(2):356-367. [Medline: 16436857]
5. Eifel PJ, Jhingran A, Bodurka DC, Levenback C, Thames H. Correlation of smoking history and other patient characteristics with major complications of pelvic radiation therapy for cervical cancer. J Clin Oncol 2002 Sep;20(17):3651-3657. [doi: 10.1200/JCO.2002.10.128] [Medline: 12202666]
6. Ditre JW, Brandon TH, Zale EL, Meagher MM. Pain, nicotine, and smoking: research findings and mechanistic considerations. Psychological Bulletin 2011;137(6):1065-1093. [doi: 10.1037/a0025544] [Medline: 21967450]
7. Seretny M, Currie GL, Sena ES, Ramnarine S, Grant R, MacLeod MR, et al. Incidence, prevalence, and predictors of chemotherapy-induced peripheral neuropathy: a systematic review and meta-analysis. Pain 2014;155(12):2461-2470. [doi: 10.1016/j.pain.2014.09.020] [Medline: 25261162]
8. Browman G, Wong G, Hodson I, Sathya J, Russell R, McAlpine L, et al. Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. N Engl J Med 1993 Jan 21;328(3):159-163. [doi: 10.1056/NEJM199301213280302] [Medline: 8417381]
9. Hamilton M, Wolf J, Rusk J, Beard SE, Clark GM, Witt K, et al. Effects of smoking on the pharmacokinetics of erlotinib. Clin Cancer Res 2006 Apr 01;12(7 Pt 1):2166-2171 [FREE Full text] [doi: 10.1158/1078-0432.CCR-05-2235] [Medline: 16609030]
10. van der Bol JM, Mathijssen RH, Loos WJ, Friberg LE, van Schaik RH, de Jonge MJ, et al. Cigarette smoking and irinotecan treatment: pharmacokinetic interaction and effects on neutropenia. J Clin Oncol 2007 Jul;25(19):2719-2726. [doi: 10.1200/JCO.2006.09.6115]

11. Baser S, Shannon VR, Eapen GA, Jimenez CA, Onn A, Lin E, et al. Smoking cessation after diagnosis of lung cancer is associated with a beneficial effect on performance status. Chest 2006 Dec;130(6):1784-1790. [doi: 10.1378/chest.130.6.1784] [Medline: 17166997]
12. Browman GP, Mohide EA, Willan A, Hodson I, Wong G, Grimard L, et al. Association between smoking during radiotherapy and prognosis in head and neck cancer: a follow-up study. Head Neck-J Sci Spec 2002 Nov 26;24(12):1031-1037. [doi: 10.1002/hed.10168] [Medline: 12454940]
13. Sitas F, Weber MF, Egger S, Yap S, Chiew M, O'Connell D. Smoking cessation after cancer. J Clin Oncol 2014 Nov 10;32(32):3593-3595. [doi: 10.1200/JCO.2014.55.9666] [Medline: 25267760]
14. Parsons A, Daley A, Begh R, Aveyard P. Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: systematic review of observational studies with meta-analysis. Br Med J 2010 Jan 21;340:b5569-b5569. [doi: 10.1136/bmj.b5569] [Medline: 20093278]
15. Mayer DK, Carlson J. Smoking patterns in cancer survivors. Nicotine Tob Res 2011 Jan;13(1):34-40. [doi: 10.1093/ntr/ntq199] [Medline: 21097514]
16. Bellizzi KM, Rowland JH, Jeffery DD, McNeel T. Health behaviors of cancer survivors: examining opportunities for cancer control intervention. J Clin Oncol 2005 Dec 01;23(34):8884-8893. [doi: 10.1200/JCO.2005.02.2343] [Medline: 16314649]
17. Hanna N, Mulshine J, Wollins D, Tyne C, Dresler C. Tobacco cessation and control a decade later: American society of clinical oncology policy statement update. J Clin Oncol 2013 Sep 01;31(25):3147-3157. [doi: 10.1200/JCO.2013.48.8932] [Medline: 23897958]
18. Shields PG, Herbst RS, Arenberg D, Benowitz NL, Bierut L, Luckart JB, et al. Smoking Cessation, Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2016 Dec;14(11):1430-1468. [doi: 10.6004/jnccn.2016.0152] [Medline: 27799513]
19. Canadian Action Network for the Advancement, Dissemination and Adoption of Practice-informed Tobacco Treatment (CAN-ADAPTT). Canadian Smoking Cessation Clinical Practice Guideline. Toronto, Canada: Centre for Addiction and Mental Health; 2011. URL: https://www.nicotinedependenceclinic.com/english/canadaptt/guideline/introduction.aspx [accessed 2019-02-10] [WebCite Cache ID 764sn1OSk]
20. Evans W, Truscott R, Cameron E, Peter A, Reid R, Selby P, et al. Lessons learned implementing a province-wide smoking cessation initiative in Ontario's cancer centres. Curr Oncol 2017 Jun;24(3):e185-e190 [FREE Full text] [doi: 10.3747/co.23.3506] [Medline: 28680285]
21. Balduyck B, Sardari Nia P, Cogen A, Dockx Y, Lauwers P, Hendriks J, et al. The effect of smoking cessation on quality of life after lung cancer surgery. Eur J Cardiothorac Surg 2011 Dec;40(6):1432-1437. [doi: 10.1016/j.ejcts.2011.03.004] [Medline: 21498082]
22. Garces YI, Yang P, Parkinson J, Zhao X, Wampfler J, Ebbert J, et al. The relationship between cigarette smoking and quality of life after lung cancer diagnosis. Chest 2004 Dec;126(6):1733-1741. [doi: 10.1378/chest.126.6.1733] [Medline: 15596667]
23. Gritz ER, Carmack CL, de Moor C, Coscarelli A, Schacherer CW, Meyers EG, et al. First year after head and neck cancer: quality of life. J Clin Oncol 1999 Jan;17(1):352-360. [doi: 10.1200/JCO.1999.17.1.352] [Medline: 10458254]
24. Jensen K, Jensen AB, Grau C. Smoking has a negative impact upon health related quality of life after treatment for head and neck cancer. Oral Oncol 2007 Feb;43(2):187-192. [doi: 10.1016/j.oraloncology.2006.02.006] [Medline: 16860590]
25. Westmaas JL, Newton CC, Stevens VL, Flanders WD, Gapstur SM, Jacobs EJ. Does a recent cancer diagnosis predict smoking cessation? An analysis from a large prospective US cohort. J Clin Oncol 2015 May 20;33(15):1647-1652. [doi: 10.1200/JCO.2014.58.3088] [Medline: 25897151]
26. Ostroff JS, Jacobsen PB, Moadel AB, Spiro RH, Shah JP, Strong E, et al. Prevalence and predictors of continued tobacco use after treatment of patients with head and neck cancer. Cancer 1995 Jan 15;75(2):569-576. [doi: 10.1002/1097-0142] [Medline: 7812925]
27. Gritz ER, Fingeret MC, Vidrine DJ, Lazev AB, Mehta NV, Reece GP. Successes and failures of the teachable moment: smoking cessation in cancer patients. Cancer 2006 Jan 01;106(1):17-27 [FREE Full text] [doi: 10.1002/cncr.21598] [Medline: 16311986]
28. Ostroff JS, Goffin JR, Khuri FR, Warren GW. Perspective on the National Comprehensive Cancer Network's Clinical Practice Guidelines for Smoking Cessation. J Oncol Prac 2016 Jan;12(1):55-58. [doi: 10.1200/JOP.2015.006148] [Medline: 26374860]
29. Warren GW, Ward KD. Integration of tobacco cessation services into multidisciplinary lung cancer care: rationale, state of the art, and future directions. Transl Lung Cancer Res 2015 Aug;4(4):339-352 [FREE Full text] [doi: 10.3978/j.issn.2218-6751.2015.07.15] [Medline: 26380175]
30. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016 Dec;25(12):986-992 [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
31. Logan J, Graham ID. Toward a comprehensive interdisciplinary model of health care research use. Sci Commun 2016 Aug 18;20(2):227-246. [doi: 10.1177/1075547098020002004]

32. Eng L, Su J, Qiu X, Palepu PR, Hon H, Fadhel E, et al. Second-hand smoke as a predictor of smoking cessation among lung cancer survivors. J Clin Oncol 2014 Feb 20;32(6):564-570. [doi: 10.1200/JCO.2013.50.9695] [Medline: 24419133]
33. Zhou W, Heist R, Liu G, Asomaning K, Miller DP, Neuberg DS, et al. Second hand smoke exposure and survival in early-stage non-small-cell lung cancer patients. Clin Cancer Res 2006 Dec 01;12(23):7187-7193 [FREE Full text] [doi: 10.1158/1078-0432.CCR-06-1460] [Medline: 17145845]
34. Zhou W, Heist RS, Liu G, Park S, Neuberg D, Asomaning K, et al. Smoking cessation before diagnosis and survival in early stage non-small cell lung cancer patients. Lung Cancer 2006 Sep;53(3):375-380. [doi: 10.1016/j.lungcan.2006.05.017] [Medline: 16814423]
35. Kashigar A, Habbous S, Eng L, Irish B, Bissada E, Irish J, et al. Social environment, secondary smoking exposure, and smoking cessation among head and neck cancer patients. Cancer 2013 Aug 01;119(15):2701-2709 [FREE Full text] [doi: 10.1002/cncr.28088] [Medline: 23765604]
36. Selby P, Brosky G, Oh P, Raymond V, Arteaga C, Ranger S. A pragmatic, randomized, controlled study evaluating the impact of access to smoking cessation pharmacotherapy coverage on the proportion of successful quitters in a Canadian population of smokers motivated to quit (ACCESSATION). BMC Public Health 2014 May 07;14:433 [FREE Full text] [doi: 10.1186/1471-2458-14-433] [Medline: 24885542]
37. Selby P, Brosky G, Oh PI, Raymond V, Ranger S. How pragmatic or explanatory is the randomized, controlled trial? The application and enhancement of the PRECIS tool to the evaluation of a smoking cessation trial. BMC Med Res Methodol 2012 Jul 23;12:101 [FREE Full text] [doi: 10.1186/1471-2288-12-101] [Medline: 22824225]
38. Costello MJ, Sproule B, Victor JC, Leatherdale ST, Zawertailo L, Selby P. Effectiveness of pharmacist counseling combined with nicotine replacement therapy: a pragmatic randomized trial with 6,987 smokers. Cancer Causes Control 2010 Dec 12;22(2):167-180. [doi: 10.1007/s10552-010-9672-9] [Medline: 21153694]
39. Herie M, Connolly H, Voci S, Dragonetti R, Selby P. Changing practitioner behavior and building capacity in tobacco cessation treatment: The TEACH project. Patient Educ Couns 2012 Jan;86(1):49-56. [doi: 10.1016/j.pec.2011.04.018] [Medline: 21612884]
40. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. Br Med J 2007 Jul 7;335(7609):24-27 [FREE Full text] [doi: 10.1136/bmj.39246.581169.80] [Medline: 17615222]
41. Stacey D, Hill S. Patient-direct and patient-mediated KT interventions. In: Straus S, Tetroe J, Graham ID, editors. Knowledge Translation in Health Care: Moving from Evidence to Practice. New Jersey: John Wiley & Sons, Ltd; Aug 01, 2013:197-211.
42. Pathman D, Konrad T, Freed G, Freeman V, Koch G. The awareness-to-adherence model of the steps to clinical guideline compliance. The case of pediatric vaccine recommendations. Med Care 1996 Sep;34(9):873-889. [Medline: 8792778]
43. 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]
44. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Acad Pediatr 2013;13(6 Suppl):S38-S44. [doi: 10.1016/j.acap.2013.08.002] [Medline: 24268083]
45. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002 Aug;27(4):299-309. [doi: 10.1046/j.1365-2710.2002.00430.x] [Medline: 12174032]
46. SAS Institute Inc. SAS/ETS 14.1 Users Guide. Cary, USA: SAS Institute Inc; 2015. Chapter 9 The AUTOREG procedure URL: http://support.sas.com/documentation/cdl/en/etsug/68148/HTML/default/viewer.htm [accessed 2019-02-10] [WebCite Cache ID 764vbq9G7]
47. Franke TM, Ho T, Christie CA. The chi-square test: often used and more often misinterpreted. Am J Eval 2011 Nov 08;33(3):448-458. [doi: 10.1177/1098214011426594]
48. Farley A, Koshiaris C, Oke J, Ryan R, Szatkowski L, Stevens R, et al. Physician support of smoking cessation after diagnosis of lung, bladder, or upper aerodigestive tract cancer. Ann Fam Med 2017 Sep 11;15(5):443-450. [doi: 10.1370/afm.2100] [Medline: 28893814]
49. Goldstein AO, Ripley-Moffitt CE, Pathman DE, Patsakham KM. Tobacco use treatment at the US National Cancer Institute's designated cancer centers. Nicotine Tob Res 2012 Apr 11;15(1):52-58. [doi: 10.1093/ntr/nts083] [Medline: 22499079]
50. Cooley ME, Emmons KM, Haddad R, Wang Q, Posner M, Bueno R, et al. Patient-reported receipt of and interest in smoking-cessation interventions after a diagnosis of cancer. Cancer 2011 Jan 10;117(13):2961-2969. [doi: 10.1002/cncr.25828] [Medline: 21692055]
51. 2008 PHS Guideline Update Panel, Liaisons,Staff. Treating tobacco use and dependence: 2008 update US Public Health Service Clinical Practice Guideline executive summary. Respir Care 2008 Sep;53(9):1217-1222 [FREE Full text] [Medline: 18807274]
52. Stead LF, Buitrago D, Preciado N, Sanchez G, Hartmann-Boyce J, Lancaster T. Physician advice for smoking cessation. Cochrane Database Syst Rev 2013 May 31(5):CD000165. [doi: 10.1002/14651858.CD000165.pub4] [Medline: 23728631]
53. Palmieri JJ, Stern TA. Lies in the doctor-patient relationship. Prim Care Companion J Clin Psychiatry 2009;11(4):163-168 [FREE Full text] [doi: 10.4088/PCC.09r00780] [Medline: 19750068]

Abbreviations

CAN-ADAPTT: Canadian Action Network for the Advancement, Dissemination and Adoption of Practice-informed Tobacco Treatment
CCO: Cancer Care Ontario
CEASE: Cessation e-referral System
HCPs: health care providers
OMRU: Ottawa Model of Research Use
PM: Princess Margaret Cancer Centre

Edited by G Eysenbach; submitted 31.07.18; peer-reviewed by J Burris, N Lisha, C Lucchiari, A Mohammadabadi; comments to author 13.09.18; revised version received 20.12.18; accepted 31.12.18; published 09.04.19.

Please cite as:
Giuliani ME, Liu G, Xu W, Dirlea M, Selby P, Papadakos J, Abdelmutti N, Yang D, Eng L, Goldstein DP, Jones JM
Implementation of a Novel Electronic Patient-Directed Smoking Cessation Platform for Cancer Patients: Interrupted Time Series Analysis
J Med Internet Res 2019;21(4):e11735
URL: https://www.jmir.org/2019/4/e11735/
doi:10.2196/11735
PMID:30964445

Original Paper

Predicting Outcomes from Engagement With Specific Components of an Internet-Based Physical Activity Intervention With Financial Incentives: Process Analysis of a Cluster Randomized Controlled Trial

Jennifer M Murray[1], BSc, PhD; David P French[2], BSc, MSc, PhD; Christopher C Patterson[1], BSc, MSc, PhD; Frank Kee[1], MSc, MD; Aisling Gough[1], BSc, PhD; Jianjun Tang[3], BS, MS, PhD; Ruth F Hunter[1], BSc, PhD

[1]Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom
[2]Manchester Centre of Health Psychology, School of Health Sciences, University of Manchester, Manchester, United Kingdom
[3]School of Agricultural Economics and Rural Development, Renmin University of China, Beijing, China

Corresponding Author:
Jennifer M Murray, BSc, PhD
Centre for Public Health
School of Medicine, Dentistry and Biomedical Sciences
Queen's University Belfast
Institute of Clinical Science B
Royal Victoria Hospital, Grosvenor Road
Belfast, BT12 6BJ
United Kingdom
Phone: 44 2890978955
Email: Jennifer.Murray@qub.ac.uk

Abstract

Background: Investigating participant engagement and nonusage attrition can help identify the likely *active ingredients* of electronic health interventions. Research on engagement can identify which intervention components predict health outcomes. Research on nonusage attrition is important to make recommendations for retaining participants in future studies.

Objective: This study aimed to investigate engagement and nonusage attrition in the Physical Activity Loyalty (PAL) scheme, a 6-month complex physical activity intervention in workplaces in Northern Ireland. The intervention included financial incentives with reward redemption and self-regulation techniques. Specific objectives were (1) to determine whether engagement in specific intervention components predicted physical activity at 6 months, (2) to determine whether engagement in specific intervention components predicted targeted mediators at 6 months, and (3) to investigate predictors of nonusage attrition for participants recording daily activity via the PAL scheme physical activity monitoring system and logging onto the website.

Methods: Physical activity was assessed at baseline and 6 months using pedometers (Yamax Digiwalker CW-701, Japan). Markers of engagement and website use, monitoring system use, and reward redemption were collected throughout the scheme. Random-effects generalized least-squares regressions determined whether engagement with specific intervention components predicted 6-month physical activity and mediators. Cox proportional hazards regressions were used to investigate predictors of nonusage attrition (days until first 2-week lapse).

Results: A multivariable generalized least-squares regression model (n=230) showed that the frequency of hits on the website's monitoring and feedback component (regression coefficient [*b*]=50.2; SE=24.5; *P*=.04) and the percentage of earned points redeemed for financial incentives (*b*=9.1; SE=3.3; *P*=.005) were positively related to 6-month pedometer steps per day. The frequency of hits on the discussion forum (*b*=−69.3; SE=26.6; *P*=.009) was negatively related to 6-month pedometer steps per day. Reward redemption was not related to levels of more internal forms of motivation. Multivariable Cox proportional hazards regression models identified several baseline predictors associated with nonusage attrition. These included identified regulation (hazard ratio [HR] 0.88, 95% CI 0.81-0.97), recovery self-efficacy (HR 0.88, 95% CI 0.80-0.98), and perceived workplace environment safety (HR 1.07, 95% CI 1.02-1.11) for using the physical activity monitoring system. The EuroQoL health index (HR 0.33, 95% CI 0.12-0.91), financial motivation (HR 0.93, 95% CI 0.87-0.99), and perceived availability of physical activity opportunities in the workplace environment (HR 0.96, 95% CI 0.93-0.99) were associated with website nonusage attrition.

Conclusions: Our results provide evidence opposing one of the main hypotheses of self-determination theory by showing that financial rewards are not necessarily associated with decreases in more internal forms of motivation when offered as part of a complex multicomponent intervention. Identifying baseline predictors of nonusage attrition can help researchers to develop strategies to ensure maximum intervention adherence.

Trial Registration: ISRCTN Registry ISRCTN17975376; http://www.isrctn.com/ISRCTN17975376 (Archived by WebCite at http://www.webcitation.org/76VGZsZug)

(J Med Internet Res 2019;21(4):e11394) doi:10.2196/11394

KEYWORDS

physical activity; workplace; randomized controlled trial; behavior; maintenance; motivation

Introduction

Background

The worldwide *pandemic* of physical inactivity [1] requires innovative approaches to increasing population physical activity levels with a view to achieving long-term maintenance [2]. Physical activity interventions that can be delivered through less costly channels (eg, internet, telephone, or post) than those requiring direct contact, in the interest of reaching as many participants as possible, are needed [3].

An example of an internet-delivered intervention was the Physical Activity Loyalty (PAL) scheme, implemented in workplaces in Northern Ireland (NI). The PAL scheme was a complex physical activity intervention that offered financial incentives and other behavior change techniques delivered via the study website to increase workplace physical activity (recorded by outdoor sensors located within 2 km of the workplace). Paradoxically, results showed that there was a small but significant decline in pedometer steps per day at 6 months relative to the baseline for the intervention group compared with controls, which dissipated at 12 months [4]. Mediation and moderation analyses showed that decreases in physical activity were partially mitigated by positive indirect effects through the constructs of integrated regulation, intrinsic motivation, and habit measured at 6 months, whereas the negative intervention effect was moderated by participants' perceptions of availability of physical activity opportunities in the workplace environment [5]. The analyses reported in this paper aimed to provide further insight to the mechanisms of behavior change for participants in the PAL scheme by examining usage rates for specific intervention components, predictors of usage rates, and whether usage was related to study outcomes.

The concept of *engagement* may be defined in terms of the level of exposure to and use of an intervention and the amount of skills practice involved (ie, completing activities or exercises to acquire knowledge or learn behavior relevant to the target outcome) [6]. A participant's level of engagement determines the extent to which they receive the intended intervention, and research on engagement is useful for identifying which intervention components are associated with health outcomes [7,8]. Investigating engagement in the different components of the intervention separately may help uncover which aspects of the intervention were beneficial (or detrimental) for increasing physical activity behavior. Thomson et al, for example, examined the intervention engagement indicators (both singly and combined) in relation to several health behaviors in a broad lifestyle intervention and recommended the use of single engagement indicators, relevant to each intervention component, for predicting health outcomes [8]. Examining engagement in this way may help identify the key *active ingredients* [9,10] for refinement in future studies. Other authors have noted that although previous studies have focused on the comparative effectiveness of Web-based interventions, they have neglected to test hypotheses about the mechanisms of action [11]. Understanding *how* and *why* interventions affect outcomes will enable the development of more efficient Web-based interventions [11]. This paper contributes to filling a gap in the research base by investigating the relationship between intervention engagement and mediator outcomes targeted by the PAL intervention. Thus, it reflects guidance provided by the Medical Research Council on conducting process evaluations, which promotes the understanding of *cause* as a key feature [12].

An issue observed to impact Web-based interventions is the tendency for a substantial proportion of participants to discontinue use of the intervention before the intervention ends [13]. Nonusage attrition refers to the phenomenon of participants ceasing intervention use before the end of the intervention period, which seems to particularly affect Web-based interventions [13]. For example, previous Web-based physical activity interventions targeting healthy, sedentary adults define nonusage attrition as occurring when the participant has a 2-week lapse from using the intervention [14,15]. To determine how successful Web-based interventions are for achieving health behavior change, it is important to understand participants' nonusage patterns and their influencing factors. Thus, we might be able to make recommendations for participant retention in future intervention studies. This is important given that lack of participant engagement and high levels of nonusage attrition are factors that can impede researchers' ability to appropriately test hypotheses in intervention studies [8,16].

Objectives

The objectives of this paper were (1) to determine whether levels of engagement in different components of the intervention predicted physical activity measured 6 months post baseline for participants assigned to the intervention group, (2) to determine whether levels of engagement in different components of the intervention predicted psychosocial variables (ie, mediators) targeted by the intervention at 6 months post baseline, and (3) to investigate rates of nonusage attrition for participants recording daily activity via the PAL scheme physical activity monitoring system and logging onto the PAL scheme website and baseline predictors of nonusage attrition (ie, sociodemographic, mediator, environmental, and physical activity variables) for participants in the intervention group.

Methods

Overview

The PAL scheme was a cluster randomized controlled trial of a complex, 6-month multicomponent workplace intervention targeting inactive employees in workplaces in Belfast and Lisburn city centers in NI to increase their physical activity during working hours [17]. The underpinning theoretical framework was based on the learning theory [18], self-regulation control theory [19], social cognitive theory [20], and self-determination theory [21]. The scheme included a novel physical activity tracking system (with sensors in outdoor locations within 2 km of the workplace) and Web-based monitoring (ie, self-monitoring, prompts and cues, habit formation, and adding objects to the environment). The main intervention component was the provision of financial incentives [22] with *points* accumulated depending on participants' minutes of walking (ie, 1 *point* for 1 min of physical activity with a notional monetary value of £0.03 for a maximum of 30 min per day) and could be redeemed for rewards at local businesses. Maps of walking routes and examples of physical activity opportunities were provided on the website (ie, instruction on how to perform the behavior). Sensors were operational during working hours (ie, 7 am-7 pm, Monday-Friday). Other behavior change techniques included regular tailored motivational emails (ie, prompts), tailored feedback, and links to other resources (eg, physical activity and healthy eating advice) [23]. Discussion forums on the website provided a platform for participants to contact researchers and other participants (ie, social support). Participants randomly assigned (in clusters) to the control arm received no intervention during the 6-month intervention period but were placed on a waiting list to participate in the scheme at the end of the study period (ie, 12 months). A more detailed overview of the trial procedures, including the Consolidated Standards of Reporting Trials (CONSORT) flow diagram, and intervention program has been published [4] and is summarized in Multimedia Appendix 1. The Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth (CONSORT-EHEALTH) checklist has been completed for this study [24].

Data Collection

Outcome data were collected at baseline (sociodemographic, mediator, environmental, and physical activity variables), 6 months (mediator and physical activity variables), and 12 months (physical activity only). Data on daily physical activity captured via the PAL scheme physical activity monitoring system (ie, dates and minutes), website usage (ie, dates, number of hits, and minutes), and reward redemption (ie, number of earned points and proportion redeemed) were collected throughout the 6-month intervention period.

Outcome Measurements

Engagement Variables

A total of 3 markers of overall intervention engagement (ie, daily physical activity captured via the PAL scheme physical activity monitoring system, use of the PAL website, and reward redemption) were tracked throughout the 6-month intervention period and the following variables were derived:

1. *Percentage of intervention days during which participants walked for at least 10 min captured via the PAL scheme physical activity monitoring system over the 6-month intervention period.* This captured participants' engagement with the physical activity monitoring system component of the intervention (ie, their willingness to practice physical activity behavior in the workplace and earn points to incentivize their physical activity). Government recommendations suggest that adults (aged 18-65 years) should accumulate 150 mins per week of moderate-intensity physical activity or 75 mins per week of vigorous-intensity physical activity, or an equivalent combination of both, in blocks of at least 10-min duration [25]. Recommendations from the Chief Medical Office emphasize the importance of daily physical activity and suggest the accumulation of 30 min of at least moderate-intensity physical activity on most, preferably all, days of the week [26]. Therefore, engagement was measured in terms of days, and only days with at least 10 min of recorded activity were counted.
2. *Percentage of intervention weeks during which participants logged onto the PAL website at least once over the 6-month intervention period.* Although there was no specific guidance for *intended use* of the PAL website (ie, there was no recommendation for how often participants should log on), research shows that the typical Web-based intervention is meant to be used once a week [27], and previous studies have categorized a log-in frequency of once per week as being high [28]. Therefore, engagement was measured in terms of weeks, and only weeks during which participants logged in at least once were counted.
3. *Percentage of earned points redeemed over the 6-month intervention period.* Aside from earning points by recording activity via the physical activity monitoring system, this indicator captured whether participants were interested in redeeming their earned points for financial rewards to incentivize their physical activity behavior.

Engagement with the different aspects of the PAL website was assessed as the *frequency of hits on each intervention component for every 10 days the participant accessed the website and the total number of intervention components accessed on the website at least once (range 0-6).* Research shows that measures of time spent on the study website may not accurately capture engagement with the intervention. For example, although Web-based interventions with unstructured access facilitate tailoring and flexibility, they enable users to multitask by opening multiple Web pages or undertaking other activities, complicating the measurement of intervention engagement [29]. As participants in the PAL scheme had complete freedom to choose how they used the website, a higher frequency of hits on a particular website component across the days on which the participant chose to log on was expected to reflect a higher level of interest (and willingness to engage) in that aspect of the intervention. This was also in line with previous studies of engagement in Web-based interventions using log-in frequency as a measure of engagement [28,30-34].

The 6 intervention components participants could access on the website were as follows:

1. *Monitoring and feedback:* Data and visual representation (ie, graphics) of the participant's activity over the intervention period for self-monitoring purposes (ie, self-monitoring and feedback and goal setting)
2. *Rewards:* Platform for participants to view their earned and bonus points, information on available rewards, and how to redeem points (ie, immediate reward contingent on behavior change)
3. *Maps:* Maps of sensor locations and example walking routes for planning of physical activity (ie, information on when and where to perform physical activity and action planning)
4. *Health information (physical activity):* Physical activity facts and information, health benefits, safety tips, and tips for a physically active lifestyle (ie, provision of information about health benefits of physical activity)
5. *Health information (other):* Information related to healthy eating, smoking, alcohol consumption, and stress reduction (ie, provision of information about health benefits of other health behaviors)
6. *Discussion forums:* Platform for participants to contact researchers and other participants to ask questions, make enquiries, raise concerns, and respond to comments (ie, social support).

Nonusage Attrition: Recording Activity Via the Physical Activity Monitoring System and Website Use

Nonusage attrition was considered to occur if a participant had at least a 2-week lapse from use [14,15]. Nonusage attrition for recording activity via the physical activity monitoring system was measured as the number of days until the first 2-week lapse from recording activity. Website nonusage attrition was measured as the number of days until the first 2-week lapse from logging onto the website.

Physical Activity

The primary outcome was steps per day objectively measured over 7 days using sealed pedometers (Yamax Digiwalker CW-701, Japan) [35-37] and considered valid if the participant provided more than or equal to 250 steps per day for 3 or more days. This was collected at baseline, 6 months, and 12 months. The primary outcome assessment was distinct from the data collected from the PAL physical activity monitoring system. Specifically, the physical activity monitoring system was used to capture data on the minutes of workplace physical activity undertaken in the outdoor workplace environment by intervention group participants within the core hours of 7 am to 7 pm, Monday to Friday. These data were used to compute participants' points, were redeemable for financial incentives, and were available on the study website as a self-monitoring tool. The primary outcome assessment was conducted by asking participants in both the intervention and control groups to wear sealed pedometers for 7 days at baseline, 6 months, and 12 months (between waking in the morning and going to bed at night, except during water-based activities).

Mediator Outcomes

Mediator outcomes were collected at baseline and 6 months via a self-reported questionnaire and included planning [38], self-determined motivation (ie, identified regulation, integrated regulation, and intrinsic motivation) [39,40], habit [41], recovery and maintenance self-efficacy [42], outcome satisfaction [43,44], social norms [45], and workplace norms [45]. These constructs were measured as they are central to the behavior change theories upon which the intervention was designed and represented the assumed pathways through which the intervention was hypothesized to lead to a change in physical activity behavior [17]. Exploring whether engagement with the various components of the intervention was related to changes in these constructs is useful to identify the *active ingredients* of the intervention, determine the degree to which the intervention worked as intended, and improve our understanding of how the intervention led to a change in behavior. Self-reported questionnaire data were collected via the Web-based platform Qualtrics (Qualtrics, Provo, Utah, USA).

Predictors of Nonusage Attrition

Predictors of nonusage attrition were sociodemographic, mediator, and environmental variables (assessed by questionnaire) and physical activity measures (pedometer steps per day) collected at baseline. Sociodemographic variables included age, gender, highest educational level, income, marital status, and self-reported height and weight (used to compute body mass index). Measures of health included Short Form-8 physical and mental health component scores [46], the Quality of Life health state utility measure and weighted health index [47], and the Warwick-Edinburgh Mental Wellbeing Scale [48,49]. Mediator variables included outcome expectations [43], physical activity self-efficacy [50], intention [51], planning [38], financial motivation [52,53], self-determined motivation (ie, identified regulation, integrated regulation, and intrinsic motivation) [39,40], habit [41], recovery and maintenance self-efficacy [42], outcome satisfaction [43,44], and social norms and workplace norms [45]. Perceptions of workplace environment (attractiveness, safety, accessibility, and availability) were also collected at baseline [54]. Descriptions of assessed variables are provided in Multimedia Appendix 2.

Statistical Analysis

These analyses are exploratory and should be interpreted with caution because of multiple testing. The level of significance was $P<.05$ for all analyses. Analyses were carried out using Stata 13 (StataCorp) [55]. All questionnaire items were coded so that higher numerical variables equaled higher values of the construct.

Objective 1: To Determine Whether Levels of Engagement in Different Components of the Intervention Predicted Physical Activity Measured 6 Months Post

Baseline for Participants Assigned to the Intervention Group

Random-effects generalized least-squares regressions were run with 6-month physical activity (ie, pedometer steps per day) as the dependent variable and engagement variables (ie, percentage of intervention days in which participants undertook at least 10 min of physical activity captured using the PAL scheme physical activity monitoring system, percentage of intervention weeks participants logged onto the PAL website, percentage of earned points redeemed, frequency of hits on each of the 6 website intervention components for every 10 days the participant accessed the website, and total number of website sections accessed at least once) as the independent variables. The model was adjusted for randomization stratum (large>50, medium=20-50, small<20 or schools or colleges), season (6-month follow-up occurred between December 2015 and April 2016 versus 6-month follow-up occurred between July 2016 and August 2016), and baseline pedometer steps per day with SEs and *P* values adjusted for clustering (3 clusters based on size and 1 cluster for educational establishments). Random-effects models explicitly modeled the dependence between observations within the same cluster by including the random effect. This represented the amount by which the intercept for a given cluster differed from the overall mean intercept value [56]. These analyses were conducted using Stata's *xtreg* command with the *vce (cluster)* option specified. Engagement variables showing a significant relationship with 6-month physical activity in univariable analyses (*P*<.05) were included in a multivariable model with backward elimination of the predictor with the highest *P* value until all included predictors had *P*<.05. This determined the combined effects of all relevant predictors on 6-month physical activity. The distributions of residuals for each regression were plotted to check for normality. Partial regression plots were used to identify influential points, and homogeneity of variances was checked by graphing residual versus fitted values.

This paper focused on pedometer steps per day collected at 6 months only as the primary study outcomes were collected at 6 months, the intervention period was 6 months, and the significant negative intervention effect observed for pedometer steps per day had dissipated at 12 months [4].

Objective 2: To Determine Whether Levels of Engagement in Different Components of the Intervention Predicted Psychosocial Variables (ie, Mediators) Targeted by the Intervention at 6 Months Post Baseline

Random-effects generalized least-squares regressions were run with 6-month mediators as the dependent variable and engagement variables (ie, percentage of intervention days in which participants undertook at least 10 min of physical activity captured using the PAL scheme physical activity monitoring system, percentage of intervention weeks participants logged onto the PAL website, percentage of earned points redeemed, frequency of hits on each of the 6 website intervention components for every 10 days the participant accessed the website, and total number of website sections accessed at least once) as the independent variables. These analyses used the same procedures outlined under Objective 1 and additionally included baseline values of the relevant mediator as a covariate.

Objective 3: To Investigate Rates of Nonusage Attrition for Participants Recording Daily Activity via the Physical Activity Loyalty Scheme Physical Activity Monitoring System and Logging onto the Physical Activity Loyalty Scheme Website, and Baseline Predictors (ie, Sociodemographic, Mediator, Environmental, and Physical Activity Variables) of Nonusage Attrition for Participants in the Intervention Group

Survival curves for time to nonusage attrition were plotted separately for participants' use of the physical activity monitoring system to record daily activity and website use. The median usage (ie, the time by which 50.0% of participants' usage had lapsed; 211/422 for use of the physical activity monitoring system to record daily activity and 209/418 for logging onto the website) was then calculated. Baseline measures of sociodemographic variables, mediator variables, environmental variables, and physical activity were investigated as predictors of nonusage attrition of the physical activity monitoring system to record daily activity and nonusage attrition for use of the website using Cox proportional hazards regression analyses. In the first analysis, the time variable was the number of days until the first 2-week lapse from using the physical activity monitoring system to record daily activity. In the second analysis, the time variable was the number of days until the first 2-week lapse from logging onto the website. For each model, the event variable was coded 1 if nonusage attrition occurred or 0 if nonusage attrition did not occur. Univariable analyses were conducted on all predictor variables and those with *P*<.05 were included in a multivariable model with backward elimination of the predictor with the highest *P* value until all included predictors had *P*<.05. All analyses included SEs and *P* values corrected for clustering. The Efron procedure was used for handling ties as it is advocated over the Breslow method [57] and can be implemented with models adjusting SEs and *P* values for clustering. The proportional hazards assumption was tested for each model formally using the Schoenfeld residuals (*P*<.05 provided evidence to reject the proportional hazards assumption), and by visual inspection of scaled Schoenfeld residual plots [58]. Plots of −log(−log[survival]) versus log(time) were created for categorical predictors with nonparallelism indicating violation of the proportional hazards assumption.

As a sensitivity analysis for our definition of nonusage attrition, we repeated these analyses defining nonusage attrition as occurring if a participant had a 1-month (ie, 30 days) lapse from use.

Results

Baseline Characteristics

A total of 457 participants from 19 clusters were randomized to the intervention group. Baseline characteristics are reported in Multimedia Appendix 2.

Engagement, Physical Activity, and Mediator Outcomes at 6 Months

Table 1 shows the 6-month engagement and nonusage attrition measures. The mean percentage of intervention days during which participants were recorded being active via the physical activity monitoring system was 24.7% (SD 21.8%; approximately 44/180 days), and the mean number of intervention weeks that participants logged onto the study website was 37.8% (SD 32.5%; approximately 9/24 weeks). Participants redeemed 39.3% (SD 42.5%; approximately 39 points for every 100 points earned) of their earned points on average. Participants clicked on 4 of the 6 website components at least once on average, and the component accessed with the highest frequency was monitoring and feedback. The 6-month physical activity and mediator outcomes are reported in Table 2.

Table 1. Descriptive statistics for 6-month engagement and nonusage attrition.

Variables	Statistics (6 months)	
	n	Mean (SD)
Engagement		
Percentage of intervention days participants walked for at least 10 min captured via the physical activity monitoring system[a]	422	24.7 (21.8)
Percentage of intervention weeks participants logged onto the website[b]	418	37.8 (32.5)
Percentage of earned points redeemed[c]	422	39.3 (42.5)
Frequency: Monitoring and feedback[d]	418	13.7 (3.5)
Frequency: Rewards[d]	418	5.7 (4.5)
Frequency: Maps[d]	418	3.4 (4.0)
Frequency: Health information (physical activity)[d]	418	0.5 (1.7)
Frequency: Health information (other)[d]	418	1.2 (3.2)
Frequency: Discussion forums[d]	418	1.9 (4.2)
Total number of sections (website)[e]	418	3.9 (1.5)
Total minutes (recording daily activity via physical activity monitoring system)	422	1000 (987)
Total minutes (PAL[f] website)	418	1171 (2048)
Nonusage attrition		
Days to nonusage attrition (recording daily activity via physical activity monitoring system)[g]	422	53.7 (61.2)
Days to nonusage attrition (PAL website)[h]	418	31.7 (43.4)
Number of participants with nonusage attrition for recording daily activity via physical activity monitoring system, n (%)	—[i]	375 (88.9)
Number of participants with PAL website nonusage attrition, n (%)	—	403 (96.4)

[a] Percentage of days participants were recorded walking for at least 10 mins captured via the physical activity monitoring system.

[b] Percentage of weeks participants logged onto the website at least once.

[c] Percentage of total accumulated points which the participant had redeemed by 6 months.

[d] Frequency of hits (ie, total number of hits for every 10 days the participant accessed the website).

[e] Number of sections accessed on website at least once (0-6).

[f] PAL: Physical Activity Loyalty.

[g] Number of days until first 2-week lapse from recording daily activity via physical activity monitoring system.

[h] Number of days until first 2-week lapse from logging onto the website.

[i] Not applicable.

Table 2. Baseline and 6-month physical activity outcomes and scores on mediator variables.

Variables (scale range)	n	Baseline, mean (SD)	n	6 Months, mean (SD)
Physical activity self-efficacy (1-5)	439	2.91 (0.97)	—[a]	—
Intentions (1-7)	435	5.38 (1.68)	—	—
Outcome expectations (1-5)	418	3.37 (0.62)	—	—
Financial motivation (1-7)	439	1.71 (1.16)	—	—
Planning (1-4)	414	2.37 (0.69)	255	2.35 (0.74)
Social norms (1-7)	414	3.87 (1.20)	253	3.90 (1.13)
Identified regulation (1-5)	438	3.81 (0.87)	262	3.93 (0.82)
Integrated regulation (1-5)	439	3.12 (1.13)	258	3.41 (1.10)
Intrinsic motivation (1-5)	438	3.52 (0.99)	259	3.70 (0.91)
Habit (1-5)	437	2.89 (1.32)	256	3.18 (1.40)
Workplace norms (1-5)	439	3.20 (0.82)	260	3.19 (0.76)
Recovery self-efficacy (1-4)	438	2.36 (0.82)	261	2.41 (0.73)
Maintenance self-efficacy (1-4)	438	2.79 (0.86)	262	2.69 (0.83)
Outcome satisfaction (1-5)	404	3.85 (0.68)	257	3.87 (0.62)
Pedometer steps per day[b]	414	7977 (3602)	249	6990 (3078)

[a] Variable not measured at 6 months.

[b] 12-month pedometer steps per day (mean 7790, SD 3462; n=210).

Objective 1: To Determine Whether Levels of Engagement in Different Components of the Intervention Predicted Physical Activity Measured 6 Months Post Baseline for Participants Assigned to the Intervention Group

Table 3 shows the results of random-effects regressions with pedometer steps per day at 6 months as the dependent variable and use of specific intervention components as the independent variable, controlling for baseline pedometer steps per day, stratum, and season, with cluster-adjusted SEs and *P* values. Engagement variables that were significant predictors of 6-month pedometer steps per day in univariable analyses were included in a multivariable model that showed that the frequency of hits on the monitoring and feedback component of the website across the 6-month intervention period (*b*=50.2; SE=24.5; *P*=.04) and percentage of earned points redeemed across the 6-month intervention period (*b*=9.1; SE=3.3; *P*=.005) were positively related to 6-month pedometer steps per day, whereas the frequency of hits on the discussion forum component of the website across the 6-month intervention period (*b*=−69.3; SE=26.6; *P*=.009) was negatively related to 6-month pedometer steps per day. None of the other variables were significant predictors of 6-month pedometer steps per day in univariable analyses.

Table 3. Results of random-effects regressions with 6-month pedometer steps per day as the dependent variable and engagement indicators as independent variables among intervention group participants providing 6-month pedometer readings. Results are adjusted for stratum, season, and baseline pedometer steps per day with cluster-adjusted standard errors and *P* values

Engagement variables	Univariable models			Multivariable model[a]		
	n	*b* (SE)	*P* value[b]	n	*b* (SE)	*P* value[b]
Engagement indicators						
Percentage of intervention days participants walked for at least 10 min captured via the physical activity monitoring system[c]	231	4.2 (8.5)	.62	—[d]	—	—
Percentage of intervention weeks participants logged onto the website[e]	234	4.4 (6.0)	.47	—	—	—
Percentage of earned points redeemed[f]	231	8.3 (4.1)	*.04*	230	9.1 (3.3)	*.005*
Website sections						
Monitoring and feedback[g]	234	66.3 (18.5)	*<.001*	230	50.2 (24.5)	*.04*
Rewards[g]	234	13.9 (36.0)	.70	—	—	—
Maps[g]	234	−46.9 (43.7)	.28	—	—	—
Health information: Physical activity[g]	234	34.9 (160.0)	.83	—	—	—
Health information: Other[g]	234	25.2 (65.9)	.70	—	—	—
Discussion forums[g]	234	−77.4 (27.1)	*.004*	230	−69.3 (26.6)	*.009*
Number of sections[h]	234	−32.4 (117.4)	.78	—	—	—

[a] *R*-squared=0.54 for multivariable model. *R*-squared=0.51 for model including covariates only (ie, stratum, season, and baseline pedometer steps per day). Empty cells in this column show variables which were not included in the multivariable model.

[b] *P* values reported in italics show statistically significant results ($P<.05$).

[c] Percentage of days participants were recorded walking for at least 10 min captured via the physical activity monitoring system.

[d] Not applicable.

[e] Percentage of weeks participants logged onto the website at least once.

[f] Percentage of total accumulated points that the participant had redeemed by 6 months.

[g] Frequency of hits (ie, total number of hits for every 10 days the participant accessed the website).

[h] Number of sections accessed on website at least once (0-6).

Objective 2: To Determine Whether Levels of Engagement in Different Components of the Intervention Predicted Psychosocial Variables (ie, Mediators) Targeted by the Intervention at 6 Months Post Baseline

The only mediator variable for which more than 1 independent variable was retained in the multivariable analysis was integrated regulation (Table 4). Engagement variables that were significant predictors of 6-month integrated regulation were included in a multivariable model that showed that the percentage of intervention days during which participants walked for at least 10 min captured via the PAL scheme physical activity monitoring system over the 6-month intervention period (b=0.008; SE=0.002; P<.001) and the frequency of hits on the monitoring and feedback component of the website across the 6-month intervention period (b=0.03; SE=0.01; P=.02) were positively related to 6-month pedometer steps per day, whereas the frequency of hits on the discussion forum component of the website across the 6-month intervention period (b=−0.02; SE=0.01; P=.02) was negatively related to 6-month pedometer steps per day. The results of all univariable analyses are presented in Multimedia Appendix 3.

Table 4. Results of multivariable random-effects regressions with 6-month integrated regulation as the dependent variable and engagement indicators as independent variables among intervention group participants providing 6-month data. Results are adjusted for stratum, season, baseline pedometer steps per day, and integrated regulation with cluster-adjusted standard errors and *P* values

Engagement variables	Univariable models			Multivariable model[a]		
	n	*b* (SE)	*P* value[b]	n	*b* (SE)	*P* value[b]
Engagement indicators						
Percentage of intervention days participants walked for at least 10 mins captured via the physical activity monitoring system[c]	238	0.007 (0.002)	*.004*	236	0.008 (0.002)	*<.001*
Percentage of intervention weeks participants logged onto the website[d]	240	0.004 (0.002)	*.02*	—[e]	—	—
Percentage of earned points redeemed[f]	—	0.000 (0.001)	.82	—	—	—
Website sections						
Monitoring and feedback[g]	240	0.03 (0.01)	*.02*	236	0.03 (0.01)	*.02*
Rewards[g]	—	0.00 (0.01)	.97	—	—	—
Maps[g]	—	−0.02 (0.01)	.10	—	—	—
Health information: Physical activity[g]	240	0.11 (0.05)	*.03*	—	—	—
Health information: Other[g]	—	0.06 (0.06)	.27	—	—	—
Discussion forums[g]	240	−0.02 (0.01)	*.03*	236	−0.02 (0.01)	*.02*
Number of sections[h]	240	0.09 (0.03)	*.005*	—	—	—

[a] *R*-squared=0.59 for multivariable model. *R*-squared=0.50 for model including covariates only (ie, stratum, season, baseline pedometer steps per day, and baseline integrated regulation). Empty cells in this column show variables which were not included in the multivariable model.

[b] *P* values reported in italics show statistically significant results (*P*<.05).

[c] Percentage of days participants were recorded walking for at least 10 mins captured via the physical activity monitoring system.

[d] Percentage of weeks participants logged onto the website at least once.

[e] Not applicable.

[f] Percentage of total accumulated points that the participant had redeemed by 6 months.

[g] Frequency of hits (ie, total number of hits for every 10 days the participant accessed the website).

[h] Number of sections accessed on website at least once (0-6).

Objective 3: To Investigate rates of Nonusage Attrition for Participants Recording Daily Activity Via the Physical Activity Loyalty Scheme Physical Activity Monitoring System and Logging onto the Physical Activity Loyalty Scheme Website, and Baseline Predictors (ie, Sociodemographic, Mediator, Environmental, and Physical Activity Variables) of Nonusage Attrition for Participants in the Intervention Group

The median usage (ie, the time by which 50.0% of participants' usage had lapsed) was 26 days for use of the physical activity monitoring system to record daily activity (nonusage attrition occurred for 211/422 participants; Figure 1) and 13 days for use of the website (nonusage attrition occurred for 209/418 participants; Figure 2). Nonusage attrition of the physical activity monitoring system to record daily activity occurred for 88.9% of participants (375/422), and website nonusage attrition occurred for 96.4% of participants (403/418). In both figures, the vertical section of the curve indicates that there was a proportion of participants who did not use the intervention component within the first 2 weeks of the intervention period (approximately equal to 25.0%, or 106/422, of the intervention group for use of the physical activity monitoring system and approximately equal to 20.0%, or 84/418, of the intervention group for use of the website).

Univariable and multivariable Cox regression analyses are presented in Multimedia Appendix 4. The multivariable analysis for use of the physical activity monitoring system to record daily activity showed that having higher levels of identified regulation at baseline (hazard ratio [HR] 0.88, 95% CI 0.81-0.97) and having higher levels of recovery self-efficacy at baseline (HR 0.88, 95% CI 0.80-0.98) reduced the risk of attrition. In contrast, having a higher perception of the safety of the workplace environment for physical activity at baseline (HR 1.07, 95% CI 1.02-1.11) was associated with a higher risk of attrition. The multivariable analysis for website use showed that having higher values on the EuroQoL weighted health index (HR 0.33, 95% CI 0.12-0.91), having higher levels of financial motivation at baseline (HR 0.93, 95% CI 0.87-0.99), or having a higher perception of the availability of physical activity opportunities in the workplace environment at baseline (HR 0.96, 95% CI 0.93-0.99) reduced the risk of attrition. Formal tests and visual inspection of plots showed no evidence for violation of the proportional-hazards assumption for the multivariable models. The results of the sensitivity analysis for our definition of nonusage attrition repeating these analyses with nonusage attrition defined as occurring if a participant had a 1-month (ie, 30 days) lapse from use have been reported in Multimedia Appendix 5.

Figure 1. Survival curve for time to nonusage attrition for recording daily activity via the physical activity monitoring system (n=422).

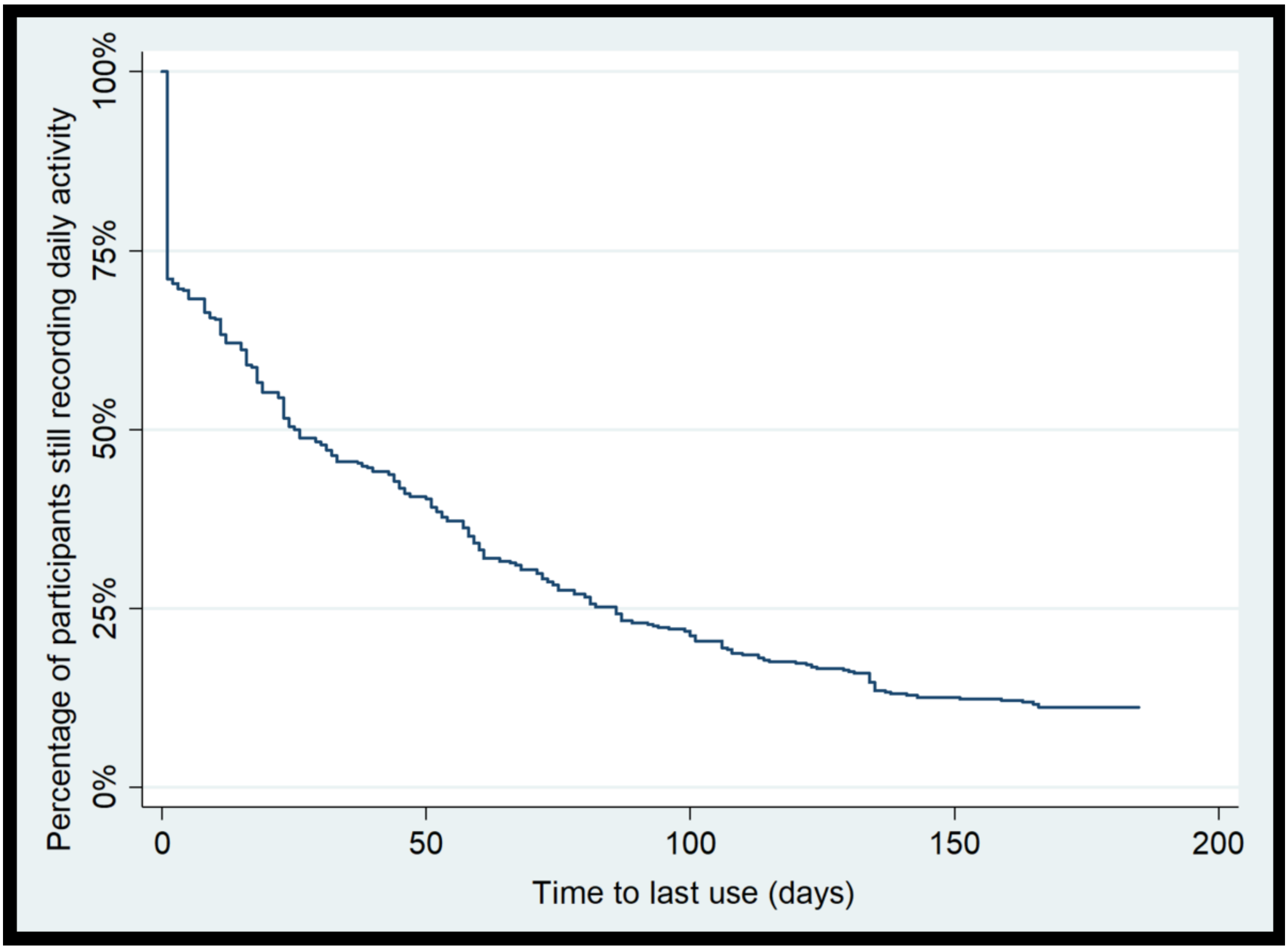
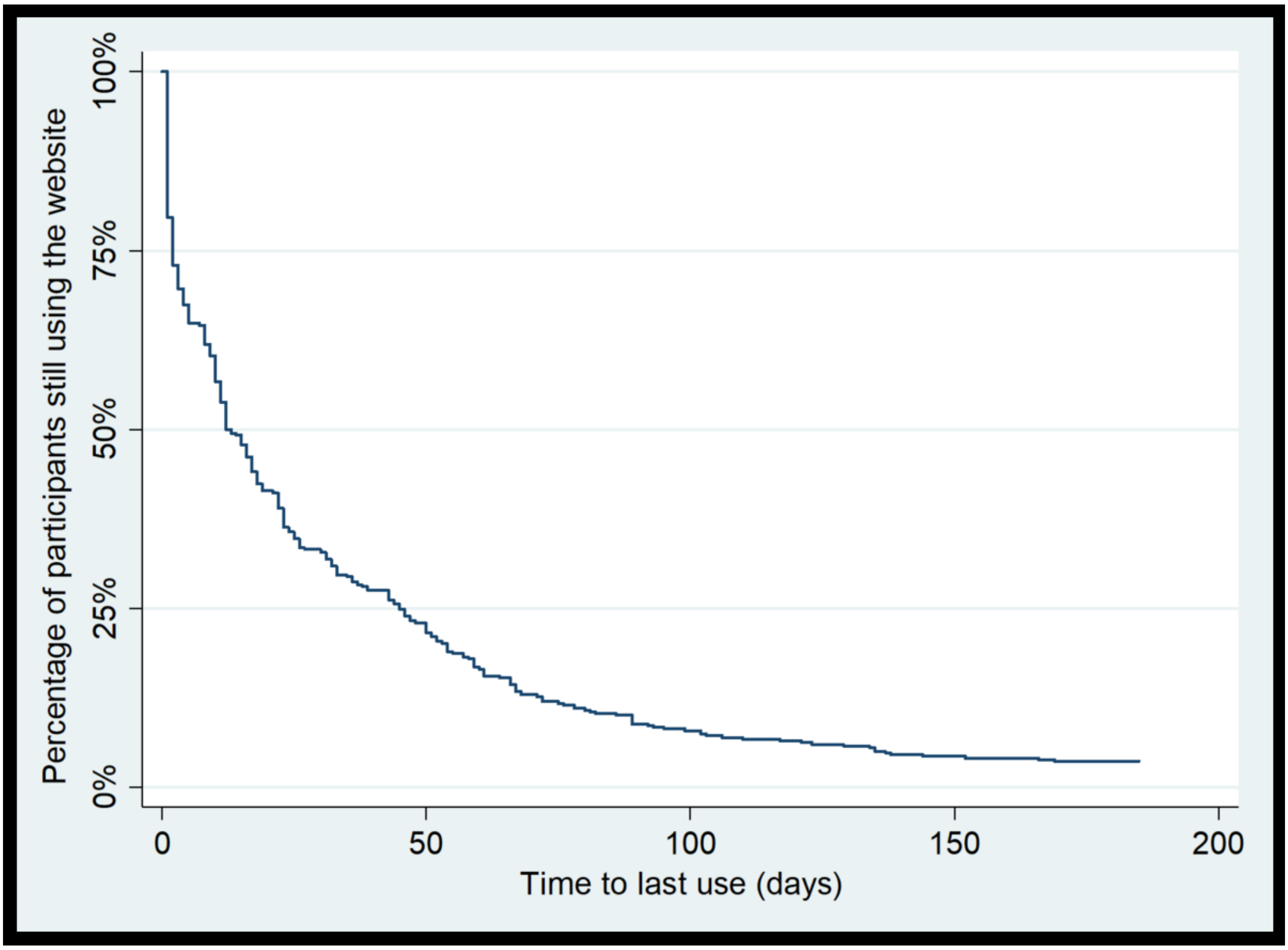


Figure 2. Survival curve for time to nonusage attrition for use of the website (n=418).



Discussion

Principal Findings

This study examined whether overall engagement (ie, using the physical activity monitoring system to record physical activity, accessing the study website, and redeeming earned points for financial rewards) in a 6-month workplace physical activity intervention (ie, the PAL scheme) and engagement with specific intervention components (ie, specific sections of the website) were associated with physical activity and mediator outcomes at 6 months. Time to nonusage attrition for different intervention components and predictors of nonusage attrition (ie, sociodemographic, mediator, environmental, and physical activity variables) were also investigated. Due to the nature of the wait-list control, there are no available data on intervention engagement and nonusage attrition for the control group. Therefore, we are cautious not to overinterpret the results and draw causal conclusions from these analyses. Multivariable generalized least-squares regression analyses revealed that higher levels of engagement with some intervention components were significantly related to 6-month pedometer steps per day. Several baseline predictors of nonusage attrition were also identified using the Cox proportional hazards regression analyses.

Intervention Engagement and Physical Activity

This study found that overall levels of engagement with the intervention (ie, using the physical activity monitoring system to record physical activity and accessing the study website) were not related to physical activity at 6 months in contrast to the findings of several previous studies [8,59]. A plausible explanation is that our indicators may not have sufficiently captured participants' true levels of engagement. Time spent on the website was not included as a measure of engagement because of the unstructured nature of website access, which meant that participants may have been engaging in other activities when logged on. Previous Web-based intervention studies investigating engagement include the website as the main intervention component and require participants to spend a significant amount of time on the website [14,15,29,32,33]. In contrast, the website was a mode of intervention delivery for the PAL study (the main intervention components were the financial incentive, placing sensors in an outdoor environment, and self-monitoring). Therefore, the time spent on the website is less relevant as an indicator of intervention engagement in this study than for the previous studies. Instead, we assumed that a higher frequency of hits on a particular section of the website for every 10 days of website use indicated higher levels of interest (and willingness to engage) in that aspect of the intervention. However, Baltierra et al noted that even this may be problematic as it gives no indication of whether participants are reading and comprehending the information or merely clicking on the various sections [29]. This illustrates the complexity inherent in measuring engagement in behavior change interventions and the need for a standardized approach [60,61].

Examining intervention engagement as a whole may not be sufficient to explain physical activity behavior change for this intervention, given the observed decline in physical activity for intervention participants. In particular, the use of multiple or multicomponent engagement measurements is recommended to track participant engagement in all the components of complex interventions [8]. Therefore, we also examined whether engagement with different intervention components was related to physical activity. The self-monitoring and feedback component was the most frequently accessed aspect of the website, and a higher frequency of accessing it was associated with a significant increase in physical activity at 6 months. This finding may indicate that when participants focused more on the scheme's self-monitoring and feedback aspects, this was associated with less of a decline in physical activity and is in line with the results of the study's mediation analyses, which found that planning and habit formation are important mediators [5]. Previous research also shows that self-regulation techniques and self-monitoring are useful strategies for physical activity behavior change or weight loss [36,62-67], and 1 previous systematic review of Web-based interventions also highlights that Web-based self-monitoring is a potentially effective technique [68]. Redeeming a higher proportion of earned points for rewards was associated with slightly higher physical activity levels at 6 months. This finding expands upon the study's mediation analyses, which found that financial motivation was not related to physical activity behavior at 6 months and proposed that it is possible that the participants did not find the financial incentives attractive enough to trigger behavior change in the first place [5]. It appears that when participants found the financial incentives desirable and redeemed their accumulated points, this was associated with less of a decline in physical activity. Previous studies have also shown that for financial incentives to be successful in inducing behavior change, the reward on offer must be deemed worthwhile to the individual participant. For example, monetary value [69] or type (eg, individual versus group based) [70] of reward can impact its effectiveness for behavior change.

A higher frequency of accessing the discussion forum component of the website was associated with a significant decline in physical activity at 6 months. Discussion forums were included on the PAL website as a means of providing social support for behavior change. For example, it was expected that participants would use these forums to contact researchers and interact with other participants to support behavior change. However, participants mainly used this component to make queries and report technological issues. This may indicate that participant frustration with some perceived limitations of the intervention, which were highlighted in a separate qualitative process evaluation [71], contributed to the overall negative impact on physical activity. A previous study finding negative intervention effects on physical activity behavior concluded that reduced support for the intervention over time was a contributing factor and cited similar reasons (eg, lack of variety in activities resulting in participant boredom and restrictions on the availability of time or space) [72].

Intervention Engagement and Mediator Outcomes

A higher frequency of accessing the website's feedback and monitoring component was also associated with increases in integrated regulation (in addition to physical activity behavior), further highlighting the importance of this particular intervention component. For example, there is evidence from the self-determination theory that more intrinsically motivated behavior is more likely to be maintained as it fulfills the basic psychological needs for competence, autonomy, and relatedness to others [10]. In comparison, engagement with the financial incentive component of the intervention (ie, redeeming a higher proportion of earned incentives for rewards or a higher frequency of accessing the website's rewards component) was not found to be related to the levels of identified regulation, integrated regulation, and intrinsic motivation. Thus, this study contributes unique evidence to contravene one of the main criticisms of financial incentives that is highlighted in the self-determination theory literature (ie, that the use of financial incentives should have a *crowding out* effect on intrinsic motivation for behaviors that are already internalized [73]). This paper provides further and supporting evidence for the findings of the study's mediation analysis to suggest financial incentives do not necessarily diminish more internal forms of motivation when delivered as part of a complex multicomponent behavior change intervention [5].

Nonusage Attrition

Nonusage attrition for use of the physical activity monitoring system and use of the website was high as most participants lapsed from using one or both features. High levels of attrition are commonly observed for use of Web-based interventions [27,74,75]; however, program usage is generally expected to be higher for controlled trials compared with freely accessible programs as participants are more likely to be motivated and committed to taking part in the study [13]. A 2012 systematic review of Web-based interventions [27] found that approximately half of the participants adhere to interventions. The definition of nonusage attrition (ie, occurring at the time of the first 2-week lapse from intervention use) adopted in this study may have contributed to the high levels of attrition observed. Although other studies of Web-based physical activity interventions have adopted this definition, it may be less applicable to the analysis of nonusage attrition in workplace interventions for which a 2-week lapse from intervention use may occur if a participant is on annual leave or is absent from work for 2 weeks or more. Therefore, every 2-week lapse from intervention use may not indicate that the participant had intentionally ceased intervention use.

The results of this analysis are consistent with findings in previous intervention studies, showing that participants with higher reported health status at baseline (versus lower health status) [76] have decreased risk for nonusage. It has previously been observed that Web-based interventions are frequently not successful in reaching individuals for whom health behavior change is needed most (eg, those with lower health status) [77]. Participants who were more financially motivated, who had higher levels of identified regulation, or who had higher levels of recovery self-efficacy at baseline were at lower risk for nonusage in this study. Individuals with higher levels of financial motivation may have been encouraged to continue participation in the scheme over time to continue benefitting from financial rewards. There is evidence that identified regulation and recovery self-efficacy are constructs that are important for long-term behavioral maintenance [78-82]. Identified regulation refers to behavior that is freely enacted based on the perceived value of its outcomes to the individual [83]. Individuals with higher levels of identified regulation may have been encouraged to engage in more continuous use of the intervention (ie, recording daily activity via the physical activity monitoring system) to achieve these valued outcomes (eg, improved health). Recovery self-efficacy refers to the individual's beliefs about their capability to return to physical activity following a lapse. Therefore, someone with higher levels of recovery self-efficacy has faith in their competence to regain control following a setback (ie, period of inactivity) [84]. It makes sense that individuals with higher recovery self-efficacy would experience a longer period of intervention usage before encountering their first 2-week lull because they are quicker to recover from a lapse.

Finally, perceptions of the workplace environment were shown to be related to nonusage risk in this study. As the intervention required participants to engage in physical activity in the outdoor environment of the workplace, it is plausible that their perceptions of the workplace environment with respect to physical activity may have influenced the use they made of the scheme. For example, an important component of the PAL scheme was the provision of information on opportunities for physical activity in the workplace environment, and intervention participants had access to maps on the study website marking out suggested walking routes. There is evidence that supportive social and physical environments facilitate behavior change maintenance by lowering the opportunity cost of behavior [85]. Taken together, these results indicate that it may be possible at baseline to identify those participants who are at the highest risk for nonusage attrition and to include strategies in the intervention design for retention.

Strengths and Limitations

A strength of this study is the examination of engagement as separate individual indicators related to the different components (ie, behavior change techniques) of the intervention. Furthermore, previous studies have investigated whether engagement (or adherence) is predictive of behavioral outcomes [31,33,59,86] without consideration of how they may relate to psychosocial outcomes (ie, mediators) targeted by the intervention and thought to lead to behavior change. Nonusage attrition was assessed in relation to use of more than 1 intervention component (ie, use of the physical activity monitoring system and the website), and this improves upon previous studies that typically assessed nonusage attrition in relation to website use only [76,87-91].

Although previous research on engagement in intervention studies has compared engagement between an intervention arm and a comparison arm [15], we were unable to include comparable engagement data from our control group because of the nature of the waitlist control. Therefore, our analysis is limited to intervention group participants only, in line with the approach adopted in previous similar studies [92]. Our analysis is also limited to responders at 6 months (ie, those who provided pedometer steps per day and mediator measurements). Furthermore, although measures of the frequency of hits on different sections of the website may indicate the participant's level of interest in a specific intervention component, they do not capture how well the participant processed the information. Another potential limitation is that only baseline variables were investigated as predictors of nonusage attrition. However, our goals were broadly similar to previous studies of predictors of nonusage attrition [90], the aim being to better describe the groups who will continue engaging in an intervention at enrollment. As previously discussed, our definition of nonusage attrition (ie, occurring at the time of the first 2-week lapse from intervention use) may have contributed to the high levels of nonusage attrition observed because of participants potentially taking a 2-week period of annual leave or other absence from work. Therefore, any period of nonuse may not have been indicative of an intentional lapse from using the intervention. However, provided such unintentional lapses did not occur differentially between groups, it can be reasonably assumed that the results of the survival analyses, indicating groups of participants who were at higher risk for nonuse, will not have been spuriously impacted. Future Web-based intervention studies implemented in workplace settings could improve the assessment of nonusage attrition by including a feature to capture whether the participant is present at the worksite.

Implications for Future Research

Future intervention studies should measure levels of engagement and nonusage with a view to making recommendations for retaining groups of participants who are at the highest risk for nonusage and lack of intervention engagement. This is particularly important for studies of Web-based interventions that are known to be particularly susceptible to lack of participant engagement and high nonusage attrition [13]. Better guidelines on how to measure intervention engagement are needed. For example, although commonly used markers of engagement (eg, the number of hits on certain website pages and time spent on the website) may indicate greater interest in different intervention components, they do not capture how much information is absorbed and processed. Regular knowledge quizzes may be helpful in this regard, but even these measures are problematic (eg, it is unclear whether the results can be attributed to intervention delivery, the participant's engagement, or another factor) [29]. Clearly, the idea of intervention engagement is complex and multifaceted [93,94]. Thus, its assessment should move beyond the utilization of simple metrics to incorporate user engagement patterns over time [93,95,96]. When attempting to define an intervention's *intended use* and assessing adherence or engagement, researchers should refer extensively to the assumed working mechanisms of the intervention. This will aid the standardization of the concepts of intervention adherence and engagement, which are often underdeveloped and improperly used in the current literature [61,94]. The Medical Research Council guidance on developing and evaluating complex interventions currently makes no reference to intervention engagement and nonusage attrition [97], and their process evaluation guidelines refer to engagement in a general way [12]. How to measure and analyze engagement and nonusage in complex public health interventions is a key gap in the literature. Researchers should consider developing engagement and retention strategies tailored toward specific groups of participants identified as being at risk for low engagement and high levels of nonusage attrition. Such endeavors should make use of behavior change theory and behavior change techniques within intervention trials, using similar approaches to how interventions are currently designed for changing behavior.

Researchers could make use of baseline data to identify participants who are at risk of nonengagement and nonusage attrition and design specific strategies to combat this. The findings of this study suggest that researchers should explore ways to keep those participants who are in worse health engaged with interventions for their entire duration. Interventionists should consider how to retain participants who are initially less financially motivated in studies whose main component is the offer of financial rewards. For studies of physical activity interventions requiring behavioral practice in the outdoor environment (as was the case for this study), participants' perceptions of the environment with respect to physical activity are important influencing factors that should be considered for engagement and nonusage.

Conclusions

More frequent use of the self-monitoring and feedback components of the intervention website (ie, self-monitoring and feedback and goal setting) and the redemption of a higher proportion of earned points for financial rewards (ie, immediate reward contingent on behavior change) were associated with increases in physical activity at 6 months for intervention group participants in the PAL study. Conversely, more frequent use of website discussion forums (ie, social support) was associated with decreases in physical activity at 6 months. A possible explanation for the negative association of discussion forum use with 6-month physical activity was that rather than making use of these forums to build social support for physical activity, participants generally used them as platforms to make queries or raise concerns. Therefore, it appears that the decline in physical activity behavior at 6 months for intervention group participants was due, at least in part, to participant dissatisfaction with some perceived study limitations (eg, technical glitches and limited financial rewards and physical activity opportunities for which rewards could be earned), which emerged in a separate qualitative process evaluation [71]. Levels of intrinsic motivation were not associated with the percentage of rewards redeemed or with the frequency of accessing the reward component of the website. Therefore, in contrast to the hypothesis of self-determination theory that offering financial rewards may *crowd out* intrinsic motivation, our results support that intrinsic motivation is not necessarily diminished when rewards are offered as part of a complex multicomponent intervention (eg, a higher frequency of accessing some of the website's other components was actually associated with higher intrinsic motivation). Rates of nonusage attrition were high, and survival analysis showed that participants who were in worse health at baseline were at higher risk for nonuse. Financial motivation, identified regulation, recovery self-efficacy, and perceptions of the environment were also risk factors for nonusage. Guidelines to measure engagement and improve nonusage attrition should be established and strategies incorporated into study design to ensure that participants adhere to interventions in their intended form.

Acknowledgments

This study has been funded via the National Institute for Health Research Public Health Research Programme, Award no: 12/211/82. The funding body had no role in the design of the study; collection, analysis, and interpretation of data; or in writing the manuscript. The work was undertaken under the auspices of the UK Clinical Research Collaboration Centre of Excellence for Public Health Research Northern Ireland (MC_CF023241), which is funded by the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the National Institute for Health Research (NIHR), and the Wellcome Trust. JMM was supported by a studentship funded by the Department for the Economy during the development of the manuscript. RFH is supported by Career Development Fellowship from the NIHR and acknowledges funding support from the HSC Research and Development Division. The authors would like to acknowledge the work of the wider PAL study team in leading the study, delivering the intervention, and collecting the data, with thanks to Dr Jen Badham for cleaning the data. Acknowledgment is due to the Institute of Electronics, Communications and Information Technology, Queen's University Belfast (for developing and maintaining the website and sensors), Club Marketing Services Ltd, Lisburn City Centre Management, the South Eastern Health and Social Care Trust, and Lisburn and Castlereagh City Council. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Public Health Research program, NIHR, National Health Service, or the Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of trial procedures and the Physical Activity Loyalty intervention programme.

[DOCX File, 320KB - jmir_v21i4e11394_app1.docx]

Multimedia Appendix 2

Description of assessed variables and baseline characteristics.

[DOCX File, 27KB - jmir_v21i4e11394_app2.docx]

Multimedia Appendix 3

Results of univariable random-effects regressions with individual mediators as dependent variables and engagement indicators as independent variables.

[DOCX File, 26KB - jmir_v21i4e11394_app3.docx]

Multimedia Appendix 4

Univariable and multivariable Cox regression analyses.

[DOCX File, 27KB - jmir_v21i4e11394_app4.docx]

Multimedia Appendix 5

Univariable and multivariable Cox regression analyses defining nonusage attrition as occurring at the first lapse from use of 1 month or longer.

[DOCX File, 31KB - jmir_v21i4e11394_app5.docx]

References

1. Andersen LB, Mota J, Di Pietro L. Update on the global pandemic of physical inactivity. Lancet 2016 Sep 24;388(10051):1255-1256. [doi: 10.1016/S0140-6736(16)30960-6] [Medline: 27475275]
2. Department of Health. 2010. Healthy lives, healthy people: our strategy for public health in England Internet URL: https://www.gov.uk/government/publications/healthy-lives-healthy-people-our-strategy-for-public-health-in-england [WebCite Cache ID 73MK85mm6]
3. Marcus B, Forsyth L. Motivating people to be physically active. Leeds: Human Kinetics; 2009:9780736072472.
4. Hunter R, Gough A, Murray J, Tang J, Brennan S, Chrzanowski-Smith O, et al. Do loyalty card schemes encourage physical activity? A cluster RCT. NIHR Journals Library 2019:-.
5. Hunter RF, Tully MA, Davis M, Stevenson M, Kee F. Physical activity loyalty cards for behavior change: a quasi-experimental study. Am J Prev Med 2013 Jul;45(1):56-63. [doi: 10.1016/j.amepre.2013.02.022] [Medline: 23790989]
6. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in Internet smoking cessation interventions: insights from a cross-sectional survey of. Internet Interv 2016 Sep;5:23-29 [FREE Full text] [doi: 10.1016/j.invent.2016.07.001] [Medline: 30135803]
7. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005 Apr;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
8. Thomson JL, Tussing-Humphreys LM, Goodman MH, Zoellner JM. Engagement indicators predict health changes in a lifestyle intervention. Am J Health Behav 2015 May;39(3):409-420. [doi: 10.5993/AJHB.39.3.13] [Medline: 25741685]
9. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]
10. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behaviour Change Theories Book - An Essential Resource for Researchers, Policy Makers and Practitioners. Surrey, UK: Silverback Publishing; 2014.
11. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS, Cobb NK, et al. Improving adherence to smoking cessation treatment: intervention effects in a web-based randomized trial. Nicotine Tob Res 2017 Mar 01;19(3):324-332 [FREE Full text] [doi: 10.1093/ntr/ntw282] [Medline: 27836982]
12. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
13. 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]

14. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 Steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
15. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
16. Blue CL, Black DR. Synthesis of intervention research to modify physical activity and dietary behaviors. Res Theory Nurs Pract 2005;19(1):25-61. [Medline: 15989166]
17. Hunter RF, Brennan SF, Tang J, Smith OJ, Murray J, Tully MA, et al. Effectiveness and cost-effectiveness of a physical activity loyalty scheme for behaviour change maintenance: a cluster randomised controlled trial. BMC Public Health 2016 Dec 22;16:618 [FREE Full text] [doi: 10.1186/s12889-016-3244-1] [Medline: 27448663]
18. Miller N, Dollard J. Social Learning and Imitation. New Haven, CT: Yale University Press; 1941:1-269.
19. Carver C, Scheier M. Attention And Self-regulation: A Control-theory Approach To Human Behavior (springer Series In Social Psychology). New York: Springer; 1981.
20. Bandura A. Self-Efficacy: The Exercise of Control. New York: Worth Publishers; 1997.
21. Deci E, Ryan R. Intrinsic motivation and Self-determination in human behaviour. New York: Springer; 1985.
22. Haff N, Patel MS, Lim R, Zhu J, Troxel AB, Asch DA, et al. The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: a meta-analysis. Am J Health Promot 2015;29(5):314-323. [doi: 10.4278/ajhp.140714-LIT-333] [Medline: 25928816]
23. Marteau TM, Ashcroft RE, Oliver A. Using financial incentives to achieve healthy behaviour. Br Med J 2009 Apr 09;338:b1415. [doi: 10.1136/bmj.b1415] [Medline: 19359291]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. World Health Organization. Global recommendations on physical activity for health. Geneva; 2010.
26. Department Of Health. Start active, stay active: report on physical activity in the UK URL: https://www.gov.uk/government/publications/start-active-stay-active-a-report-on-physical-activity-from-the-four-home-countries-chief-medical-officers [accessed 2018-10-22] [WebCite Cache ID 73MLqaIJO]
27. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
28. van den Berg MH, Ronday HK, Peeters AJ, Voogt-van der Harst EM, Munneke M, Breedveld FC, et al. Engagement and satisfaction with an internet-based physical activity intervention in patients with rheumatoid arthritis. Rheumatology (Oxford) 2007 Mar;46(3):545-552. [doi: 10.1093/rheumatology/kel341] [Medline: 17043050]
29. 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]
30. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008 Apr;76(2):341-346. [doi: 10.1037/0022-006X.76.2.341] [Medline: 18377129]
31. Zbikowski SM, Hapgood J, Smucker Barnwell S, McAfee T. Phone and web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. J Med Internet Res 2008 Nov 14;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
32. 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]
33. 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]
34. Yu CH, Parsons JA, Mamdani M, Lebovic G, Hall S, Newton D, et al. A web-based intervention to support self-management of patients with type 2 diabetes mellitus: effect on self-efficacy, self-care and diabetes distress. BMC Med Inform Decis Mak 2014 Dec 14;14:117 [FREE Full text] [doi: 10.1186/s12911-014-0117-3] [Medline: 25495847]
35. Bassett DR, Ainsworth BE, Leggett SR, Mathien CA, Main JA, Hunter DC, et al. Accuracy of five electronic pedometers for measuring distance walked. Med Sci Sports Exerc 1996 Aug;28(8):1071-1077. [Medline: 8871919]
36. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]

37. Schneider PL, Crouter SE, Lukajic O, Bassett DR. Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. Med Sci Sports Exerc 2003 Oct;35(10):1779-1784. [doi: 10.1249/01.MSS.0000089342.96098.C4] [Medline: 14523320]
38. Sniehotta F, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul;35(4):565-576. [doi: 10.1002/ejsp.258]
39. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
40. Wilson P, Rodgers W, Loitz C, Scime G. "It's who I am...really!" The importance of integrated regulation in exercise contexts. J Appl Biobehav Res 2007;11(2):79-104. [doi: 10.1111/j.1751-9861.2006.tb00021.x]
41. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Social Pyschol 2003 Jun;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
42. Scholz U, Sniehotta F, Schwarzer R. Predicting physical exercise in cardiac rehabilitation: the role of phase-specific self-efficacy beliefs. J Sport Exerc Psychol 2005;27(2):151. [doi: 10.5167/uzh-102309]
43. Finch EA, Linde JA, Jeffery RW, Rothman AJ, King CM, Levy RL. The effects of outcome expectations and satisfaction on weight loss and maintenance: correlational and experimental analyses--a randomized trial. Health Psychol 2005 Nov;24(6):608-616. [doi: 10.1037/0278-6133.24.6.608] [Medline: 16287407]
44. Rothman AJ, Sheeran P, Wood W. Reflective and automatic processes in the initiation and maintenance of dietary change. Ann Behav Med 2009 Dec;38(Suppl 1):S4-17. [doi: 10.1007/s12160-009-9118-3] [Medline: 19787308]
45. Ball K, Jeffery RW, Abbott G, McNaughton SA, Crawford D. Is healthy behavior contagious: associations of social norms with physical activity and healthy eating. Int J Behav Nutr Phys Act 2010 Dec 07;7:86 [FREE Full text] [doi: 10.1186/1479-5868-7-86] [Medline: 21138550]
46. Ware J, Kosinski M, Dewey J, Gandek B. How to score and interpret single-item health status measures: a manual for users of the of the SF-8 health survey. Lincoln RI: QualityMetric Incorporated; 2001.
47. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
48. Lloyd K, Devine P. Psychometric properties of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. J Ment Health 2012 Jun;21(3):257-263. [doi: 10.3109/09638237.2012.670883] [Medline: 22574955]
49. 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]
50. Marcus BH, Rossi JS, Selby VC, Niaura RS, Abrams DB. The stages and processes of exercise adoption and maintenance in a worksite sample. Health Psychol 1992;11(6):386-395. [Medline: 1286658]
51. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Boston: Addison Wesley; 1975.
52. Ryan R, Frederick C, Lepes D, Rubio N, Sheldon K. Intrinsic motivation and exercise adherence. Int J Sport Psychol 1997;28(4):335-354 [FREE Full text]
53. Moller AC, McFadden HG, Hedeker D, Spring B. Financial motivation undermines maintenance in an intensive diet and activity intervention. J Obes 2012;2012:8 [FREE Full text] [doi: 10.1155/2012/740519] [Medline: 22548152]
54. Ogilvie D, Mitchell R, Mutrie N, Petticrew M, Platt S. Perceived characteristics of the environment associated with active travel: development and testing of a new scale. Int J Behav Nutr Phys Act 2008 May 30;5:32 [FREE Full text] [doi: 10.1186/1479-5868-5-32] [Medline: 18513430]
55. StataCorp. Stata. College Station, Texas (TX): StataCorp LP; 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP URL: https://www.stata.com/ [accessed 2019-03-05] [WebCite Cache ID 76dvs2EpH]
56. Stevens A. The advanced handbook of methods in evidence based healthcare. California: SAGE; 2001.
57. Borucka J. Methods for handling tied events in the Cox proportional hazard model. Studies of Oeconomica Posnaniensia 2014;2(2):91-106 [FREE Full text]
58. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982 Apr;69(1):239. [doi: 10.2307/2335876]
59. Heesch KC, Mâsse LC, Dunn AL, Frankowski RF, Mullen PD. Does adherence to a lifestyle physical activity intervention predict changes in physical activity? J Behav Med 2003 Aug;26(4):333-348. [Medline: 12921007]
60. Graham L, Wright J, Walwyn R, Russell AM, Bryant L, Farrin A, et al. Measurement of adherence in a randomised controlled trial of a complex intervention: supported self-management for adults with learning disability and type 2 diabetes. BMC Med Res Methodol 2016 Dec 06;16(1):132 [FREE Full text] [doi: 10.1186/s12874-016-0236-x] [Medline: 27716063]
61. 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]
62. Hobbs N, Godfrey A, Lara J, Errington L, Meyer TD, Rochester L, et al. Are behavioral interventions effective in increasing physical activity at 12 to 36 months in adults aged 55 to 70 years? A systematic review and meta-analysis. BMC Med 2013 Mar 19;11:75 [FREE Full text] [doi: 10.1186/1741-7015-11-75] [Medline: 23506544]

63. Abraham C, Graham-Rowe E. Are worksite interventions effective in increasing physical activity? A systematic review and meta-analysis. Health Psychol Rev 2009 Mar;3(1):108-144. [doi: 10.1080/17437190903151096]
64. Bird EL, Baker G, Mutrie N, Ogilvie D, Sahlqvist S, Powell J. Behavior change techniques used to promote walking and cycling: a systematic review. Health Psychol 2013 Aug;32(8):829-838 [FREE Full text] [doi: 10.1037/a0032078] [Medline: 23477577]
65. 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]
66. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]
67. 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. [doi: 10.1016/J.AMEPRE.2018.07.005]
68. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Dec 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
69. Finkelstein EA, Linnan LA, Tate DF, Birken BE. A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. J Occup Environ Med 2007 Sep;49(9):981-989. [doi: 10.1097/JOM.0b013e31813c6dcb] [Medline: 17848854]
70. Kullgren JT, Troxel AB, Loewenstein G, Asch DA, Norton LA, Wesby L, et al. Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. Ann Intern Med 2013 Apr 02;158(7):505-514 [FREE Full text] [doi: 10.7326/0003-4819-158-7-201304020-00002] [Medline: 23546562]
71. Gough A, Prior L, Kee F, Hunter RF. Physical activity and behaviour change: the role of distributed motivation. Crit Public Health 2018 Oct 18:1-13. [doi: 10.1080/09581596.2018.1535169]
72. Azevedo LB, Burges Watson D, Haighton C, Adams J. The effect of dance mat exergaming systems on physical activity and health-related outcomes in secondary schools: results from a natural experiment. BMC Public Health 2014 Sep 12;14:951 [FREE Full text] [doi: 10.1186/1471-2458-14-951] [Medline: 25217144]
73. Promberger M, Marteau TM. When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. Health Psychol 2013 Sep;32(9):950-957 [FREE Full text] [doi: 10.1037/a0032727] [Medline: 24001245]
74. Brindal E, Hendrie GA, Freyne J, Noakes M. Incorporating a static versus supportive mobile phone app into a partial meal replacement program with face-to-face support: randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 18;6(4):e41 [FREE Full text] [doi: 10.2196/mhealth.7796] [Medline: 29669704]
75. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Dec 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
76. Jahangiry L, Shojaeizadeh D, Montazeri A, Najafi M, Mohammad K, Yaseri M. Adherence and attrition in a web-based lifestyle intervention for people with metabolic syndrome. Iran J Public Health 2014 Sep;43(9):1248-1258 [FREE Full text] [Medline: 26175979]
77. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007 Jan 22;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
78. Deci E, Ryan R. Intrinsic motivation and self-determination in human behavior. New York: Springer; 1985.
79. Deci E, Ryan R. Facilitating optimal motivation and psychological well-being across life's domains. Can Psychol 2008;49(1):14-23. [doi: 10.1037/0708-5591.49.1.14]
80. Ryan R. University of BATH. 2009. Self-determination theory and wellbeing URL: http://www.bath.ac.uk/soc-pol/welldev/wed-new/network/research-review/Review_1_Ryan.pdf [accessed 2018-10-23] [WebCite Cache ID 73Nye6LfF]
81. Schwarzer R. Self-efficacy in the adoption and maintenance of health behaviors: theoretical approaches and a new model. In: Self-efficacy Thought Control action. Washington DC: Hemisphere Publishing Corporation; 1992:217-243.
82. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
83. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [Medline: 8169757]
84. Luszczynska A, Schwarzer R. Planning and self-efficacy in the adoption and maintenance of breast self-examination: a longitudinal study on self-regulatory cognitions. Psychol Health 2003 Jan;18(1):93-108. [doi: 10.1080/0887044021000019358]
85. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]

86. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
87. Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, et al. Effectiveness of a Web 2.0 intervention to increase physical activity in real-world settings: randomized ecological trial. J Med Internet Res 2017 Dec 13;19(11):e390 [FREE Full text] [doi: 10.2196/jmir.8484] [Medline: 29133282]
88. 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]
89. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010 Feb 10;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
90. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010 Dec 14;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
91. Lim S, Norman R, Clifton P, Noakes M. Weight loss and attrition in overweight and obese young women during a 36-week internet-based lifestyle intervention. J Obes Weight Loss Ther 2014;4(4):8. [doi: 10.4172/2165-7904.1000235]
92. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
93. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: development and implementation of the Growing Healthy App Engagement Index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
94. 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 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
95. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone-based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Dec 01;19(11):e356 [FREE Full text] [doi: 10.2196/jmir.7928] [Medline: 29092811]
96. Kim JY, Wineinger NE, Taitel M, Radin JM, Akinbosoye O, Jiang J, et al. Self-monitoring utilization patterns among individuals in an incentivized program for healthy behaviors. J Med Internet Res 2016 Dec 17;18(11):e292 [FREE Full text] [doi: 10.2196/jmir.6371] [Medline: 27856407]
97. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Medical Research Council. Developing evaluating complex interventions: new guidance Internet URL: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/ [accessed 2018-10-23] [WebCite Cache ID 73O0aDvK8]

Abbreviations

HR: hazard ratio
NI: Northern Ireland
NIHR: National Institute for Health Research
PAL: Physical Activity Loyalty

Edited by G Eysenbach; submitted 26.06.18; peer-reviewed by E Fernandes de Mello Araujo; comments to author 09.10.18; revised version received 28.11.18; accepted 06.01.19; published 19.04.19.

Please cite as:
Murray JM, French DP, Patterson CC, Kee F, Gough A, Tang J, Hunter RF
Predicting Outcomes from Engagement With Specific Components of an Internet-Based Physical Activity Intervention With Financial Incentives: Process Analysis of a Cluster Randomized Controlled Trial
J Med Internet Res 2019;21(4):e11394
URL: https://www.jmir.org/2019/4/e11394/
doi:10.2196/11394
PMID:31002304

Review

What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews

Y Alicia Hong[1,2], PhD; Chen Liang[3], PhD; Tiffany A Radcliff[2], PhD; Lisa T Wigfall[4], PhD; Richard L Street[5], PhD

[1]Department of Health Administration and Policy, George Mason University, Fairfax, VA, United States
[2]School of Public Health, Texas A&M University, College Station, TX, United States
[3]Arnold School of Public Health, University of South Carolina, Columbia, SC, United States
[4]Department of Health Kinesiology, Texas A&M University, College Station, TX, United States
[5]Department of Communication, Texas A&M University, College Station, TX, United States

Corresponding Author:
Y Alicia Hong, PhD
Department of Health Administration and Policy
George Mason University
4400 University Drive, MS 1J3
Fairfax, VA, 22030
United States
Phone: 1 703 993 1929
Email: yhong22@gmu.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2019/7/e14823/

Abstract

Background: The number of patient online reviews (PORs) has grown significantly, and PORs have played an increasingly important role in patients' choice of health care providers.

Objective: The objective of our study was to systematically review studies on PORs, summarize the major findings and study characteristics, identify literature gaps, and make recommendations for future research.

Methods: A major database search was completed in January 2019. Studies were included if they (1) focused on PORs of physicians and hospitals, (2) reported qualitative or quantitative results from analysis of PORs, and (3) peer-reviewed empirical studies. Study characteristics and major findings were synthesized using predesigned tables.

Results: A total of 63 studies (69 articles) that met the above criteria were included in the review. Most studies (n=48) were conducted in the United States, including Puerto Rico, and the remaining were from Europe, Australia, and China. Earlier studies (published before 2010) used content analysis with small sample sizes; more recent studies retrieved and analyzed larger datasets using machine learning technologies. The number of PORs ranged from fewer than 200 to over 700,000. About 90% of the studies were focused on clinicians, typically specialists such as surgeons; 27% covered health care organizations, typically hospitals; and some studied both. A majority of PORs were positive and patients' comments on their providers were favorable. Although most studies were descriptive, some compared PORs with traditional surveys of patient experience and found a high degree of correlation and some compared PORs with clinical outcomes but found a low level of correlation.

Conclusions: PORs contain valuable information that can generate insights into quality of care and patient-provider relationship, but it has not been systematically used for studies of health care quality. With the advancement of machine learning and data analysis tools, we anticipate more research on PORs based on testable hypotheses and rigorous analytic methods.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42018085057; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=85057 (Archived by WebCite at http://www.webcitation.org/76ddvTZ1C)

(J Med Internet Res 2019;21(4):e12521) doi:10.2196/12521

KEYWORDS

patient review websites; patient online review; systematic review

Introduction

People have increasingly turned to the internet to share their clinical experience and make comparisons of physicians and medical treatments [1,2]. Hundreds if not thousands of patient online reviews (PORs) appear daily on the crowdsource platforms of patient review websites (PRWs) and carry growing influence in patients' medical decision making [1-4]. In the earlier debates of PORs, some physicians expressed skepticism; they worried that most PORs were posted by begrudged patients who were not able to assess the technical quality of health care delivery [5]. Furthermore, physicians are unable to refute a negative review without jeopardizing patient confidentiality [6]; and it is nearly impossible to verify if the comments were left by actual patients [3]. Also, even with an increasing number of PORs, most rated physicians average a handful of ratings, which is unlikely to reflect the full range of impressions made by a physician who sees hundreds of patients each year [6]. Proponents of PORs, however, argue that patients are like consumers of other services and therefore have a right to express their opinions about services they pay for, and PORs provide timely and direct customer feedback [3,6,7].

Despite the ongoing debates on whether PORs can improve the quality of care [8,9], the number of PORs has grown exponentially in the past decade [1,10,11]. A recent national survey in the United States revealed that 59% of participants reported PORs were very important or somewhat important when choosing a physician, though PORs were endorsed less frequently than other factors such as word of mouth from family and friends and whether the physician accepted one's insurance [2].

The proliferation of PORs and popularity of PRWs has happened in 2 somewhat overlapping contexts. Of these, 1 is that the ubiquitous internet access has facilitated online consumer behaviors, featured by "electronic word of mouth" [12]. People go online to rate any product or service they purchase and check online ratings before making any purchase. Health care consumer behaviors, though lagging other consumer behaviors, are rapidly catching up [3]. The other context is the movement of patient empowerment and self-determination of medical care, alongside the more recognized importance of patient experience and patient satisfaction in evaluating health care quality [13,14]. For example, the Center for Medicare and Medicaid Services (CMS) has a set of Core Quality Measures for Healthcare, and "patient experience" is one of the 7 critical domains [15]. Traditional government- or health care organization (HCO)–initiated surveys have incorporated patient-reported outcome measures in their routine questionnaires of quality measures, but it takes years to conduct surveys and analyze the data, and few patients have access to or understand these data [13]. Within such contexts, PORs have become a consumer-driven alternative that can provide almost instant feedback of health care experience.

The increasing weight of PORs in patients' health care decision making has led to a growing number of research studies on PORs and PRWs [1,11]. Some scholars have advocated for giving more scientific values to PORs [7,16]. Others have evaluated the quality of PRWs and examined public perceptions and use of PRWs [2,10,17,18]. They concluded that the research on and usage of PRWs was limited [17,18]. To date, no systematic review of POR studies was available. Accordingly, we conducted a systematic review with the aims to synthesize existing studies on PORs by summarizing study characteristics, research design, analytical methods, and major findings. We have depicted the trend of POR research, identified literature gaps, and made recommendations for future research.

Methods

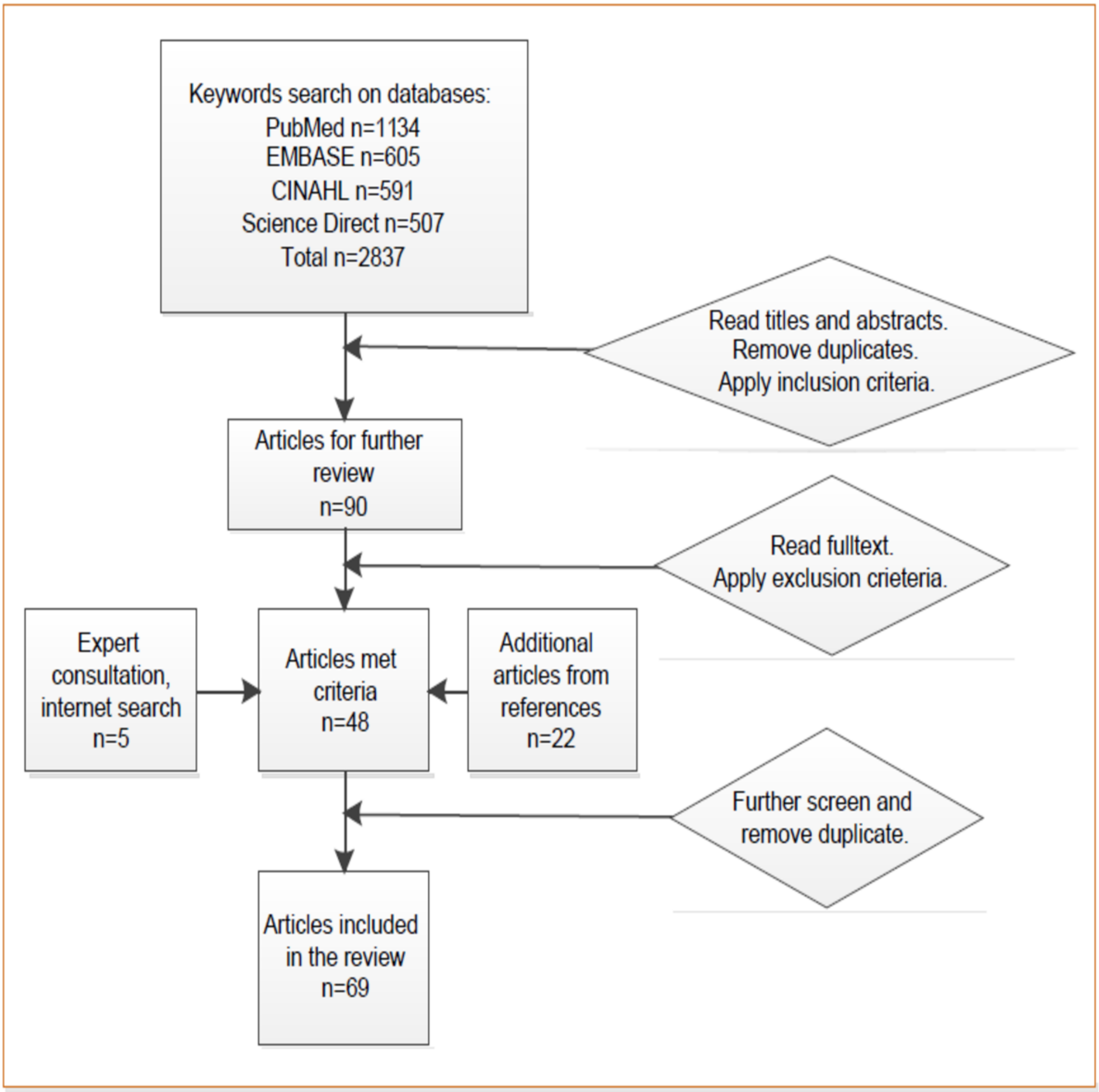
Inclusion and Exclusion Criteria

On the basis of the research objectives mentioned above, we listed the following search criteria before we started the literature search. The inclusion criteria were as follows: (1) studies that focused on PORs of physicians or hospitals, (2) studies that reported qualitative or quantitative results from analyses of PORs, and (3) peer-reviewed studies written in English. The exclusion criteria were (1) studies that did not report empirical outcomes from analyses of PORs and (2) editorials, reviews, or commentaries. Excluded studies were, for example, focused on physicians' responses to online reviews [19], reported innovative methods for analyzing PORs without reporting the analytical results [20,21], or focused on characteristics of the patients who had used PRWs without reporting POR-related outcomes [2,22].

Data Sources and Selection

Following the principles of Preferred Reporting Items for Systematic Reviews and Meta-Analyses [23], we searched the major databases of PubMed, EMBASE, CINAHL, and Science Direct in January 2019. Search terms from previously published studies were used [17,18,24], including rating sites (websites), review sites (websites), online reviews (ratings), doctor (physician and hospital) ratings, and patient reviews (ratings). As shown in Figure 1, the initial search identified 2837 articles. After reviewing the titles and abstracts to determine relevance and removing duplicates, 90 articles were further reviewed by reading full texts. A total of 48 articles that met the inclusion and exclusion criteria were identified for detailed review. Next, we searched the reference sections of the 48 articles and consulted experts in the field to identify additional articles by hand search, resulting in 26 additional articles for review. After removing duplicates, we identified 69 articles or 63 studies to include in the review. Articles that reported findings based on the same data source, similar design, and research questions were counted as one study. The systematic review protocol was registered in PROSPERO: International Prospective Register of Systematic Reviews.

Figure 1. Flow chart of the literature search and article retrieval.



Data Extraction and Synthesis

A total of 2 researchers independently reviewed all articles and extracted the following information using a predesigned table: authors and publication year, study time and location, PRWs used in the study, type of providers being studied, number of PORs and providers being analyzed, study design and analytical approach, and key findings. The intercoder reliability, calculated by using Cohen kappa, was 0.86. In the key finding analysis, the researchers listed the bullet points of major findings from each study and discussed the discrepancies until a consensus was reached. Owing to the heterogeneity of the studies, no study appraisal was carried out. This review was not focused on a single health outcome; instead, we aimed to identify and synthesize available POR studies, and no meta-analysis was conducted.

Results

Study Time and Location

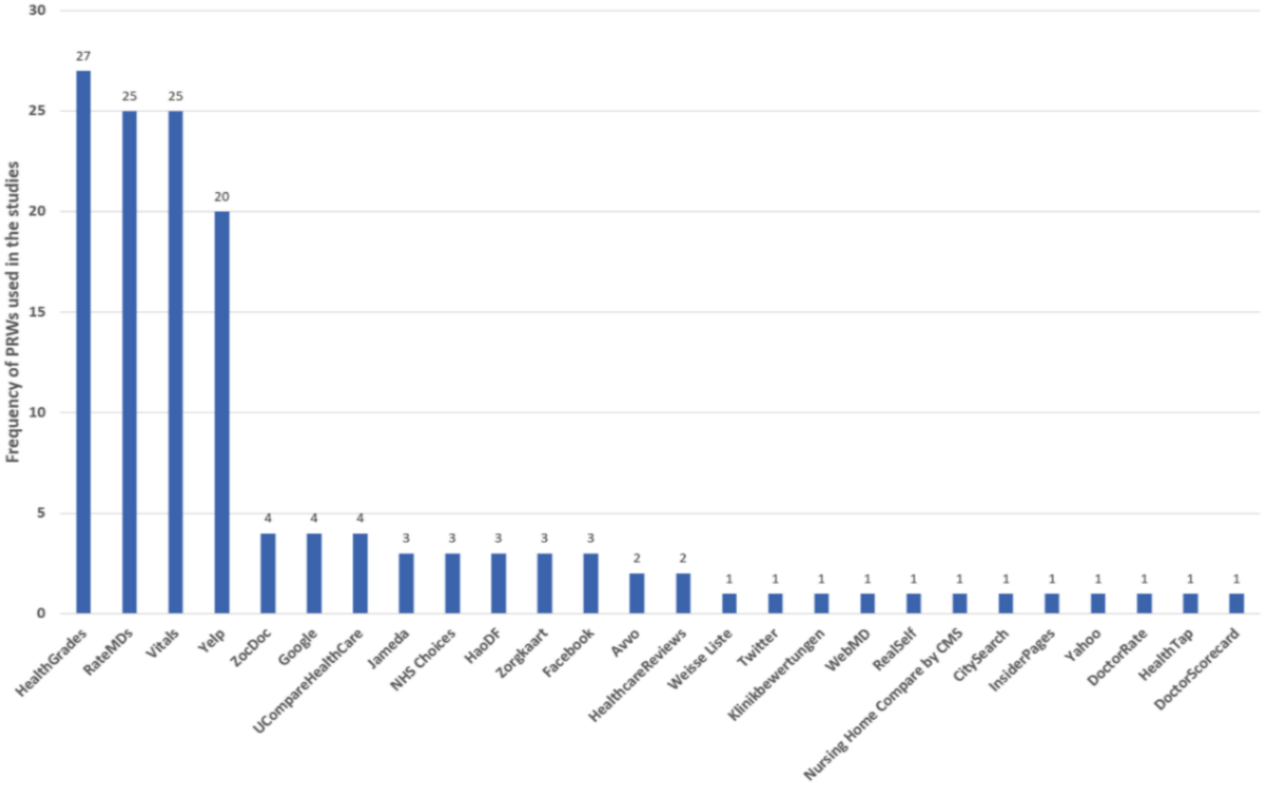
A total of 63 studies (69 articles) were included in the reviews (Multimedia Appendix 1) [1,3,10,11,25-85]. Although PRWs have been available for more than 2 decades, the earliest study on PORs was published in 2009 [25], and most of the studies (61/63, 96.8%) were published after 2010. Out of 63 studies, 48 were conducted in the United States, including Puerto Rico, 5 in Germany [26-31], 3 in the United Kingdom [32-34], 3 in China [35-38], 3 from the Netherlands [39-41], 1 from Australia [42], and 1 from Canada [86] (Table 1).

Table 1. Study locations.

Country	Statistics, n (%)
United States	48 (76.2)
Germany	5 (7.9)
United Kingdom	3 (4.8)
China	3 (4.8)
The Netherlands	3 (4.8)
Australia	1 (1.6)
Canada	1 (1.6)
Total	64[a]

[a] One study was conducted in both China and the United States.

Figure 2. Patient review websites (PRWs) used in the studies.



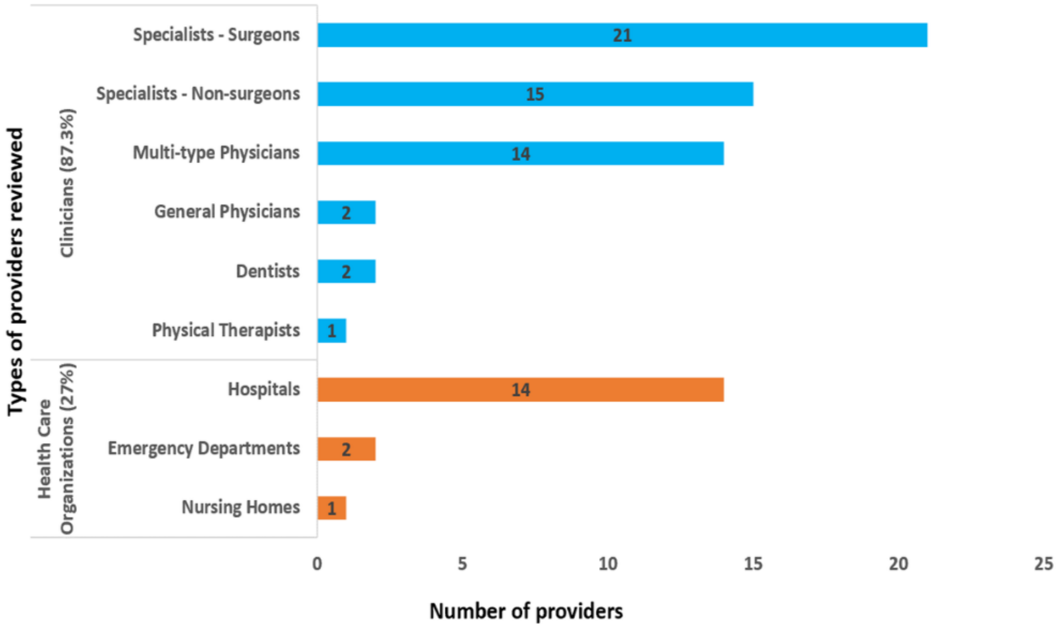
Patient Review Websites

Most studies (36/63, 57.1%) retrieved PORs from multiple PRWs with 27 studies (27/63, 42.9%) using a single PRW for data analysis. Some earlier studies googled “patient review” or providers’ names to retrieve PORs, and the most popular or promoted PRWs emerged through such an online search. The PRWs used in these research studies varied across countries. For example, in the United Kingdom, most physicians and hospitals were rated on the National Health System Choices website, which was a single PRW used in the studies from the United Kingdom [32-34]. The most popular PRW in Germany was Jameda [28-31], whereas HaoDF was used in China [35-38].

In the United States, a large number of PORs have accrued on the most popular PRWs including generic consumer review sites such as Yelp, Yahoo, and Google, as well as specialized PRWs such as RateMDs, HealthGrades, and Vitals (see Figure 2).

Studies that compared multiple PRWs found a low correlation between these sites [43,44]. For example, Nwachukwu et al reported that the correlations (*r*) between PRWs were 0.32 approximately 0.51, *P*<.001[43]. Physicians on one PRW were rated differently on other PRWs, whereas no PRW contained all consensus core domains of quality measures [45]. Some studies questioned the reliability of PRWs given that most physicians only have a very small number of ratings [10,46].

Figure 3. Various types of providers reviewed.



Types of Providers Reviewed

Out of the 63 studies, 17 (17/63, 27%) reported PORs of HCOs, including hospitals, urgent care centers (emergency departments), and nursing homes [27,32,34,39,40,44,47-53,87,88], and 55 (55/63, 87.3%) were focused on clinicians. Of the 55 studies that reported PORs of clinicians, 14 (14/55, 25.5%) included multi-type physicians (general practitioners and specialists), 2 (2/55, 3.6%) were focused on general practitioners [33,54,55], 1 (1/55, 1.8%) reported physical therapists [88], 2 (2/55, 3.6%) reported dentists [30,88], and the remaining 36 (36/55, 70.6%) were focused on specialists, including surgeons, dermatologists, urologists, Ob/Gyns. Of these 36 studies on specialists, 21 (21/36, 58.3%) were focused on various types of surgeons (see Figure 3).

Number of Providers Reviewed and Number of Ratings

The number of health care providers reviewed ranged from 20 to 212,933 with a median of 600. The number of PORs analyzed ranged from 30 to 2,685,066 with a median of 5439. The number of PORs included in the analyses has grown substantially in the past 9 years.

Not all physicians had an online rating. In Germany, only 37% of physicians were rated online [29]; a recent study in 3 metros of the United States reported that 34% of physicians did not have any PORs and most physicians still had no more than 1 review [10]. Even among those physicians with online reviews, the number of PORs varied significantly across specialties. For example, 96% of cardiologists in the United States were rated online [56]; 25% of the hospitals included in Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) had Yelp ratings [51]. Specialists were twice as likely to be rated online than general practitioners [26]; radiologists and pathologists were least likely to be rated online [42].

Study Design and Analytical Approaches

A considerable number of studies (27/63, 42.9%) were descriptive and reported only frequency analyses, including the average numbers of ratings per provider, the percentages of providers that have been reviewed online, and the mean scores of PORs. Studies that focused on HCOs and specialists typically identified the providers from a directory before searching for their PORs. In contrast, studies that focused on all types of providers typically retrieved PORs directly from PRWs without a preselected list of providers.

A total of 19 (30.2%) studies analyzed the narrative comments of PORs. Previous studies in this regard used traditional qualitative methods to retrieve major themes from these comments [3,25,32,54,55]. Recent studies have applied more advanced techniques such as natural language processing (NLP). For example, topic models, such as Latent Dirichlet Allocation, have been used as an efficient tool to automatically cluster POR comments by topics [35,36,38,50,87]. The use of such advanced analytical methods enabled content analysis of hundreds of thousands of narrative comments.

More than half of the studies (n=38/63, 60.3%) employed a comparative design. They typically compared PORs with (1) traditional surveys of patient experience [27,34], (2) providers' characteristics [28,29,33,34,57,58], (3) clinic outcomes such as patients' readmission rates or mortality rates [52,56,59], and (4) traditional "golden standards" of health care quality indicators (eg, HCAHPS structural and quality of care measures) [39,44,45,47,50,53]. Furthermore, 1 study compared PORs between China and the United States [38] and 2 studies compared PORs across different PRWs [43,44].

General Findings of Patient Online Reviews

Most patients made positive comments about their providers and would recommend their providers to families and friends. Out of 63 studies, 27 studies (27/63, 42.9%) reported mean scores of PORs ranging from 2.37 to 4.51 (out of 5) with a median value of 4 and a mean value of 3.89 (unweighted).

The studies that analyzed the patients' narrative comments found that these comments covered the entire health care encounter of the facility and staff [54,55], including physicians' demeanor, staff friendliness, empathy, and cost [60,61]; patients also cared about the ease of scheduling, time spent with patients, and wait time [62].

Relationship of Patient Online Reviews and Providers' Characteristics

The existing studies that compared PORs with characteristics of providers found that physicians with higher ratings had the following characteristics: (1) female, young age [29,43,46]; (2) more online presence [58]; (3) board-certified with extensive training experiences and graduated from a highly rated medical school [1,89]; (4) active status and years in practice [63]; (5) specialties [37,88]; and (6) locations [86]. However, some studies found no interactions between PORs and either genders, regions, or academic proclivity [46,64,65]. Furthermore, 1 study found surgeons with higher volume of procedures had higher POR ratings and better comments [59]. Patient characteristics also affected PORs. For example, female, seniors, and patients covered by private insurance were more likely to provide positive PORs [27,30].

Relationship of Patient Online Reviews and Traditional Patient Surveys

As summarized in Table 2, convergent findings suggested a strong association between PORs and traditional patient satisfaction surveys. For example, several studies found moderate-to-high degrees of correlation between PORs and HCAHPS patient experience measures [31,33,34,53,66] and the Press Ganey Medical Practice Survey for patient satisfaction [67], respectively. Content analysis studies also reported a considerable overlap between the narrative comments of PORs and thematic domains of HCAHPS surveys [47,50,51]. Some of these studies also identified additional domains not included in HCAHPS surveys. [50,51]. Similar findings of correlation tests were reported from studies in Germany [31] and the United Kingdom [33,34]. Furthermore, 2 studies from the Netherlands reported that hospitals under supervision or inspection from authorities had lower POR ratings [39,40].

Relationship of Patient Online Reviews and Clinical Outcomes and Other Quality Measures

Table 2 also includes the summaries of the relationship between PORs and clinical outcomes and other quality measures. Most of the 9 studies on the relationships of PORs and clinical outcomes reported weak or no relationship [11,31,33,52,53,56,66,68,69]. For instance, a study of PORs on cardiologists found no correlation (Spearman $\rho=-0.06$; $P=.13$) between PORs and mortality rates following the coronary artery bypass surgery [56]. Similarly, Greaves et al found a weak correlation between PORs and clinical outcomes of providers in the United Kingdom (Spearman $\rho=-0.18$ approximately 0.18; $P<.001$) [33]. By contrast, Bardach et al compared PORs of hospitals in Yelp with quality measures from HCAHPS and found that higher scores of PORs were associated with better clinical outcomes, including lower mortality and readmission rates [53]. Studies also reported significant but low degrees of association between PORs and (1) patient likelihood of visiting their primary care physicians within 14 days of discharge [11], (2) cost of care [11,66], (3) 30-day readmission and length of stay [69], and (5) other hospital level CMS quality measures [44].

Table 2. Studies that compare patient online reviews with traditional healthcare quality indicators.

Study	Comparator measures (patient surveys, clinical outcomes, or other quality measures)	Comparison methods and results
Greaves et al, 2012 [33]	(1) Mailed-based patient surveys. (2) Clinical outcomes from the National Health Service (NHS) Information Center and NHS Comparators (eg, The proportion of patients with diabetes receiving flu vaccinations, proportion of hypertensive patients with controlled blood pressure, proportion of diabetic patients with controlled HbA1C, percentage of low-cost statin prescribing, cervical screening rate, admission rates for ambulatory care sensitive conditions, and the proportion of achieved clinical Quality and Outcomes Framework (QOF) points from available points. N (POR)=16,592, N (physicians)=4934.	(1) ρ =0.37~0.48, $P<.001$ for Pearson correlation of POR and survey. (2) ρ=–0.18~0.18, $P<.001$ for the correlation of POR and clinical outcomes.
Greaves et al, 2012 [34]	Traditional survey of patient experience. N (POR)=9,9997, N (physicians)=146.	ρ=0.13~0.49, $P<.001$ for Pearson correlation of POR and survey.
Segal et al, 2012 [59]	Volume of surgeries. N of POR=588, N of surgeons=600.	High volume surgeons have higher mean values of PORs than low-volume surgeons, but effect size was weak.
Bardach et al, 2013 [53]	(1) Overall hospital ratings on HCAHPS. (2) Hospital individual HCAHPS domain scores (eg, nurse communication, pain control). (3) Hospital 30-day mortality and hospital 30-day readmission rates. N (POR)=3796, N (hospitals)=962.	Pearson correlation (n=270), ρ=0.49, $P<.001$ for 3 out of 4 measures. Higher ratings were associated with lower mortality and readmission rates.
Wallace et al, 2014 [11]	(1) Likelihood of patient visiting their primary care physician within 14 days of hospital discharge. (2) Health care expenditure. N (POR)=58,110, N (physicians)=19,636.	(1) Regression model for sentiment generated from POR comments and the comparator r^2=.21, P=.03; (2) Regression model for POR rating combined with topics generated from POR comments r^2=.25.
Glover et al, 2015 [52]	30-day hospital-wide all-cause unplanned readmission rate (HWR). POR=Facebook comments. POR=Facebook comments, N (hospitals)=136.	Independent sample t test (n=315 vs 364), POR=4.15±0.31 vs 4.05±0.41, $P<.01$ more PORs was associated lower HWR.
Emmert et al, 2015 [31]	(1) Quality measures on cost of medication, type 2 diabetes-related intermediate outcome measure, and patient/doctor ratio from German Integrated Health Care Network (QuE); (2) German patient satisfaction survey from QuE. N (POR)=1179 on Jameda, N=991 on Weisse Liste. N (physicians)=69.	(1) Spearman's rank correlation (n=991) ρ=0.297~.384, $P<.05$ for cost per prescription; ρ=0.478, $P<.05$ for patient with HbA1c-target values; ρ=−0.316~−0.289, $P<.05$ for patient/doctor ratio on Weisse Liste; (n=1179) ρ=0.298, $P<.05$ for cost per case, ρ=0.298~386, $P<.05$ for patient/doctor ratio on Jameda; (2) Spearman's rank correlation (n=991), ρ=−0.347~−0.372, $P<.05$ for 3 out of 4 measures on Weisse Liste; (n=1179), ρ=−0.391~0.640, $P<.05$ for all measures on Jameda.
Okike et al, 2016 [56]	Risk-adjusted mortality rate. N of POR NA[a], N (surgeons)=590.	Pearson's correlation (n=590), r=−.06, P=.13.
Bardach et al, 2016 [51]	Researchers identified HCAHPS domains. N (POR)=244 (narratives), N (hospitals)=193.	Content analysis (139/244, 57% of POR comments mentioned HCAHPS domains).
Kilaru et al, 2016 [47]	HCAHPS inpatient care surveys. N (POR)=1736, N (Emergency Departments)=100.	Content analysis. Considerable overlaps in theme of PORs and HCAHPS domains.
Ranard et al, 2016 [50]	Researchers identified HCAHPS domains. N (POR)=16,862, N (hospitals)=1352.	Content analysis. POR comments covered 7/11 HCAHPS domains and introduced 12 new domains not existing in HCAHPS.
Emmert et al, 2018 [44]	Hospital-level quality measures by the CMS. N (POR)=1000, N (hospitals)=623.	(1) Spearman's correlation ρ=±0.143, $P<.05$ for 13 of 29 measures; (2) Spearman's correlation ρ=±0.114, $P<0.05$ for 7 of 29 measures, indicating weak association.
Trehan et al, 2018 [68]	Total knee replacement (TKR) outcomes: infection rate, 30-day readmission rate, 90-day readmission rate, revision surgery. N of POR NA, N (surgeons)=174.	Kruskal–Wallis one-way analysis of variance one-way analysis of variance (one-way ANOVA on ranks) showed no correlation.
Campbell et al, 2018 [66]	1) HCAHPS patient satisfaction measures; 2) HCAHPS hospital-wide 30-day readmission rate; 3) Medicare spending per beneficiary ratio. N of POR NA, N (hospitals)=136.	(1) Multivariable linear regression (n=136), r^2=.16~.5, $P<.05$ for 21 of 23 measures; Pearson's correlation (n=136), r=.27~.61, $P<.005$ for 19 of 23 measures; (2) Multivariable linear regression r^2=−.58, $P<.10$ for readmission rate; (3) Multivariable linear regression r^2=−.006, $P<.731$ for Medicare spending per beneficiary. Overall weak association.
Jarari et al, 2018 [71]	Nursing Home Compare (NHC) website quality measures.	POR rating was significantly different from NHC rating.

Study	Comparator measures (patient surveys, clinical outcomes, or other quality measures)	Comparison methods and results
Chen et al, 2018 [67]	Press Ganey Medical Practice Survey for patient satisfaction. N of POR NA, N (physicians)=200.	Pearson's correlation (n=226), r=.18, *P*<.001.
Daskivich et al, 2018 [69]	Specialty-specific performance scores (adherence to Choosing Wisely measures, 30-day readmissions, length of stay, and adjusted cost of care), primary care physician peer-review scores, and administrator peer-review scores.	Multivariable linear regression (n=30) r=−.04, *P*=.04.

[a]NA: not available.

Discussion

To the best of our knowledge, this is the first systematic review of studies on PORs. The 63 studies included in this review reflect a decade of peer-reviewed publications on PORs from 6 countries; the study design and key findings have been summarized. Earlier studies tended to report on characteristics of PORs whereas later studies tended to compare PORs with traditional patient surveys or clinical outcomes.

Principal Findings

Our summaries of the existing 63 studies on PORs revealed that the number of health care providers (including clinicians and HCOs) being reviewed represented only a small number of the total health care workforce. The number of reviews per clinician varied from zero to hundreds, indicating a very skewed distribution in these PORs. As compared with general practitioners, specialists, especially surgeons, were more likely to be reviewed and included in the analyses of PORs. Overall, the online ratings and comments were positive. Only a small number of studies compared the correlations between PORs and patient satisfaction and clinical outcomes. These studies suggested that PORs were highly correlated to the "patient experience" measured by traditional patient surveys. Nevertheless, there were inconclusive findings on whether PORs were inconsistent with traditional measures of clinical outcomes. Notably, reviewed studies have identified several domains of patient experience that were not covered by the traditional patient surveys, for example, HCAHPS [50,51].

The current literature on PORs suggests a relatively new but fast-growing field. The number of published studies was small when compared with the exponential growth of PORs. Therefore, we have made the following recommendations for future studies on PORs.

First, studies with rigorous design, longitudinal nature, and larger samples are needed. POR studies present challenges of data acquisition and processing because of the nature of large and heterogeneous online data. The latest Web crawling techniques have enabled efficient retrieval of large quantities of POR data. Advanced analytical techniques such as machine learning and NLP can be employed to expedite large-scale analysis of PORs.

Second, most existing studies are focused on specialists in metropolitan areas [10,70-72]; more studies are needed to understand other disciplines of health care providers and those who serve in nonmetro areas. Studies that identify consumer-based assessments for underrepresented types of HCOs, such as nursing homes, public health services, and substance treatment centers, are minimal or missing in the literature. There was only 1 study that reported PORs for nursing homes [49]. Many of these HCOs serve vulnerable populations who are not typical PRW users, but their family caregivers and other advocates may also provide valid PORs.

Third, we anticipate more studies that go beyond the simple descriptive analysis and test theory-based hypotheses to provide more clinical and policy implications. In recent years, we have observed emerging studies that compared PORs with traditional measures of patient experience and clinical outcomes. However, the current literature is limited in terms of a lack of consistent POR reporting and insufficient advanced statistical analyses of POR data and their relationship with quality measures. We call for more empirical studies with meaningful hypotheses, rigorous design, and appropriate data analytics.

Finally, we have observed that PORs have begun clustering on a small number of popular PRWs (Figure 2). With the recent announcement of Amazon's entrance into health care [90], online reviews by health care consumers may become even more clustered. Whether and how the clustering of PORs on the growing dominance of commercial PRWs would affect consumer health behaviors and health care quality remains unstudied.

Policy Implications

The growing body of literature on PORs indicates its increasing importance in patients' decision making, which provides policy and practice implications for health care providers, patients, PRW owners, and policy makers.

Notably, health care providers should not underestimate the importance of PORs. Instead, they should recognize the importance of PRWs for their "digital brand" and stay aware of the PORs posted to popular PRWs [91]. Physicians can use anonymous PORs for the evaluation of patient satisfaction and assessment of patients' need. In addition, friendly and personalized responses to PORs may enhance positive patient-provider communication [19].

From a consumer's perspective, patients need to know that only a small number of physicians have been reviewed online and the average rating score for a physician might not be sufficient for choosing a doctor as assumed, given the tendency of consumers to provide feedback on experiences that are unusually positive or negative. As posting the health care experience becomes more commonplace, we anticipate a "consumer's guide" to help patients navigate the PORs and make more informed choices [92,93].

For PRW owners, as PORs are often unstructured, not adjusted for risks, and unverifiable, they should take more social responsibilities by adding design components to enable identity authentication, to remove inflammatory or abusive comments, and to assist patients on how to use PRWs to avoid misinformation [3,4,94]. We also call for a consistent rating scheme to facilitate the evaluation of providers using data from various PRWs.

For policy makers, the question of whether PORs can be used as an indicator of health care quality is still controversial; policy makers and health care providers should acknowledge and embrace its increasing importance for patients [7,95]. The PORs can reflect instant feedback of patients' medical encounters, the context of their ratings, and what they truly value. Some of the constructs of patient experience identified from analyzing PORs can be used to strengthen or complement the current measures of health care quality and to provide rapid recognition of quality perception gaps along with service corrections or other proactive quality interventions when needed [96]. Although we recognize the growing weight of PORs in consumer health behaviors and the potential of applying PORs in improving health care quality, we call for broader collaborations of key stakeholders, including patients, caregivers, health care providers, PRW owners, policy makers, and health services researchers, to engage in conversations and joint efforts to construct a positive patient-provider feedback loop.

Some potential biases need to be noted while interpreting the results from this review. First, this review was focused on the published studies that analyzed PORs, so the findings related to PORs only reflected those published studies but not the whole picture of PORs. Given the vast and ever-growing number of PORs, only a small fraction was studied and published. Second, only a small number of patients would actually provide ratings of their medical encounters. These motivated patients are more likely to be younger, female, living in metropolitans, and spending more time online [4]; thus, there is a potential bias in the existing PORs. These biases are not methodological flaws in conducting the systematic review but require caution when interpreting study findings.

Limitations

In addition to the above potential biases, we should also note the limitations of the study. Though we tried our best to thoroughly search the major databases, it is possible that some relevant studies were missed. As we concluded the search in January 2019, a few recently accepted papers were not included. Our search was limited to peer-reviewed literature; we may have missed some gray literature that is equally important for the POR research. Additionally, because our review was limited to the literature published in English, the review did not cover articles published in other languages. Finally, because of the heterogeneity in outcome reporting and study design, we did not carry out an appraisal of study quality. The number of PORs ranged from a few dozens to hundreds of thousands and the ratings were based on different scales, so we did not conduct a meta-analysis.

Conclusions

To conclude, the current body of the peer-reviewed literature on PORs is still small but growing rapidly. We found that overall PORs tended to be positive, and the narratives of PORs have provided insights into multiple domains of patient experience and health care quality. We call for more research on PORs using rigorous design and large samples along with better use of POR information by patients, physicians, and policy makers. We also advocate for recommendations or guidelines of POR use to help patients make informed choices and foster the application of PORs for improving health care quality.

Acknowledgments

The study was supported by the PESCA award and T3 grant from Texas A&M University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summaries of published studies on patient online reviews (63 studies consisting of 69 articles).

[PDF File (Adobe PDF File), 190KB - jmir_v21i4e12521_app1.pdf]

References

1. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
2. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. J Am Med Assoc 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
3. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]

4. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
5. Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Reuben DB, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. Ann Intern Med 2006 May 2;144(9):665-672. [doi: 10.7326/0003-4819-144-9-200605020-00010] [Medline: 16670136]
6. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]
7. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking patients' narratives about clinicians from anecdote to science. N Engl J Med 2015 Aug 13;373(7):675-679. [doi: 10.1056/NEJMsb1502361] [Medline: 26267629]
8. Bacon N. Will doctor rating sites improve standards of care? Yes. Br Med J 2009;338:b1030. [Medline: 19293223]
9. McCartney M. Will doctor rating sites improve the quality of care? No. Br Med J 2009;338:b1033. [Medline: 19293224]
10. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. J Am Med Assoc 2017 Dec 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
11. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
12. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? J Interact Market 2004 Jan;18(1):38-52. [doi: 10.1002/dir.10073]
13. Fung CH, Lim Y, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008 Jan 15;148(2):111-123. [Medline: 18195336]
14. Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. Med Care Res Rev 2010 Feb;67(1):27-37. [doi: 10.1177/1077558709341065] [Medline: 19638641]
15. Center for Medicare and Medicaid Services. Core Quality Measures URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/ [accessed 2018-10-17] [WebCite Cache ID 73ErAIAl3]
16. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. J Am Med Assoc 2010 Oct 20;304(15):1711-1712. [doi: 10.1001/jama.2010.1499] [Medline: 20959582]
17. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
18. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
19. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
20. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4] [Medline: 22207014]
21. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
22. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
23. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
24. Verhoef LM, van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
25. Black EW, Thompson LA, Saliba H, Dawson K, Black NM. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253 [FREE Full text] [Medline: 20359403]
26. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
27. Drevs F, Hinz V. Who chooses, who uses, who rates: the impact of agency on electronic word-of-mouth about hospitals stays. Health Care Manage Rev 2014;39(3):223-233. [doi: 10.1097/HMR.0b013e3182993b6a] [Medline: 23727786]

28. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
29. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
30. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating website: an analysis of the ratings. J Med Internet Res 2015;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
31. Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. BMC Health Serv Res 2015;15(1):414 [FREE Full text] [doi: 10.1186/s12913-015-1051-5] [Medline: 26404452]
32. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. Jt Comm J Qual Patient Saf 2013 Jan;39(1):7-15. [Medline: 23367647]
33. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
34. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
35. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
36. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 1;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
37. Zhang W, Deng Z, Hong Z, Evans R, Ma J, Zhang H. Unhappy patients are not alike: content analysis of the negative comments from China's Good Doctor website. J Med Internet Res 2018 Jan 25;20(1):e35 [FREE Full text] [doi: 10.2196/jmir.8223] [Medline: 29371176]
38. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: international comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
39. Kool RB, Kleefstra SM, Borghans I, Atsma F, van de Belt TH. Influence of intensified supervision by health care inspectorates on online patient ratings of hospitals: a multilevel study of more than 43,000 online ratings. J Med Internet Res 2016 Dec 15;18(7):e198 [FREE Full text] [doi: 10.2196/jmir.5884] [Medline: 27421302]
40. Hendrikx RJ, Spreeuwenberg MD, Drewes HW, Struijs JN, Ruwaard D, Baan CA. Harvesting the wisdom of the crowd: using online ratings to explore care experiences in regions. BMC Health Serv Res 2018 Oct 20;18(1):801 [FREE Full text] [doi: 10.1186/s12913-018-3566-z] [Medline: 30342518]
41. van de Belt TH, Engelen LJL, Verhoef LM, van der Weide MJ, Schoonhoven L, Kool RB. Using patient experiences on Dutch social media to supervise health care services: exploratory study. J Med Internet Res 2015;17(1):e7 [FREE Full text] [doi: 10.2196/jmir.3906] [Medline: 25592481]
42. Atkinson S. Current status of online rating of Australian doctors. Aust J Prim Health 2014;20(3):222-223. [doi: 10.1071/PY14056] [Medline: 24852125]
43. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a Sports Medicine Surgeon's. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
44. Emmert M, Meszmer N, Schlesinger M. A cross-sectional study assessing the association between online ratings and clinical quality of care measures for US hospitals: results from an observational study. BMC Health Serv Res 2018 Dec 5;18(1):82 [FREE Full text] [doi: 10.1186/s12913-018-2886-3] [Medline: 29402321]
45. Ramkumar PN, Navarro SM, Chughtai M, La T, Fisch E, Mont MA. The patient experience: an analysis of orthopedic surgeon quality on physician-rating sites. J Arthroplasty 2017 Dec;32(9):2905-2910. [doi: 10.1016/j.arth.2017.03.053] [Medline: 28455178]
46. Kirkpatrick W, Abboudi J, Kim N, Medina J, Maltenfort M, Seigerman D, et al. An assessment of online reviews of hand surgeons. Arch Bone Jt Surg 2017 May;5(3):139-144 [FREE Full text] [Medline: 28656160]
47. Kilaru AS, Meisel ZF, Paciotti B, Ha YP, Smith RJ, Ranard BL, et al. What do patients say about emergency departments in online reviews? A qualitative study. BMJ Qual Saf 2016 Jan;25(1):14-24. [doi: 10.1136/bmjqs-2015-004035] [Medline: 26208538]
48. Tran NN, Lee J. Online reviews as health data: examining the association between availability of health care services and patient star ratings exemplified by the Yelp Academic dataset. JMIR Public Health Surveill 2017 Jul 12;3(3):e43 [FREE Full text] [doi: 10.2196/publichealth.7001] [Medline: 28701293]
49. Johari K, Kellogg C, Vazquez K, Irvine K, Rahman A, Enguidanos S. Ratings game: an analysis of Nursing Home Compare and Yelp ratings. BMJ Qual Saf 2018 Aug;27(8):619-624. [doi: 10.1136/bmjqs-2017-007301] [Medline: 29133461]

50. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. Health Aff (Millwood) 2016 Apr;35(4):697-705. [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
51. Bardach NS, Lyndon A, Asteria-Peñaloza R, Goldman LE, Lin GA, Dudley RA. From the closest observers of patient care: a thematic analysis of online narrative reviews of hospitals. BMJ Qual Saf 2016 Dec;25(11):889-897 [FREE Full text] [doi: 10.1136/bmjqs-2015-004515] [Medline: 26677215]
52. Glover M, Khalilzadeh O, Choy G, Prabhakar AM, Pandharipande PV, Gazelle GS. Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance outliers. J Gen Intern Med 2015 Oct;30(10):1440-1446 [FREE Full text] [doi: 10.1007/s11606-015-3236-3] [Medline: 25749881]
53. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
54. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
55. Detz A, López A, Sarkar U. Long-term doctor-patient relationships: patient perspective from online reviews. J Med Internet Res 2013;15(7):e131 [FREE Full text] [doi: 10.2196/jmir.2552] [Medline: 23819959]
56. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
57. Riemer C, Doctor M, Dellavalle RP. Analysis of online ratings of dermatologists. JAMA Dermatol 2016 Feb;152(2):218-219. [doi: 10.1001/jamadermatol.2015.4991] [Medline: 26677133]
58. Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online patient ratings of hand surgeons. J Hand Surg Am 2016 Jan;41(1):98-103. [doi: 10.1016/j.jhsa.2015.10.006] [Medline: 26710742]
59. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
60. Smith RJ, Lipoff JB. Evaluation of dermatology practice online reviews: lessons from qualitative analysis. JAMA Dermatol 2016 Feb;152(2):153-157. [doi: 10.1001/jamadermatol.2015.3950] [Medline: 26606326]
61. Goshtasbi K, Lehrich BM, Moshtaghi O, Abouzari M, Sahyouni R, Bagheri K, et al. Patients' online perception and ratings of neurotologists. Otol Neurotol 2019 Jan;40(1):139-143. [doi: 10.1097/MAO.0000000000002075] [Medline: 30531643]
62. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: analysis of 2185 reviews. Am J Orthop (Belle Mead NJ) 2014 Aug;43(8):359-363. [Medline: 25136868]
63. Haglin JM, Eltorai AE, Kalagara S, Kingrey B, Durand WM, Aidlen JP, et al. Patient-rated trust of spine surgeons: influencing factors. Global Spine J 2018 Oct;8(7):728-732 [FREE Full text] [doi: 10.1177/2192568218767385] [Medline: 30443484]
64. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
65. Skrzypecki J, Przybek J. Physician review portals do not favor highly cited US ophthalmologists. Semin Ophthalmol 2018;33(4):547-551. [doi: 10.1080/08820538.2017.1332234] [Medline: 28628346]
66. Campbell L, Li Y. Are Facebook user ratings associated with hospital cost, quality and patient satisfaction? A cross-sectional analysis of hospitals in New York State. BMJ Qual Saf 2018 Dec;27(2):119-129. [doi: 10.1136/bmjqs-2016-006291] [Medline: 28600452]
67. Chen J, Presson A, Zhang C, Ray D, Finlayson S, Glasgow R. Online physician review websites poorly correlate to a validated metric of patient satisfaction. J Surg Res 2018 Jul;227:1-6. [doi: 10.1016/j.jss.2018.01.037] [Medline: 29804840]
68. Trehan SK, Nguyen JT, Marx R, Cross MB, Pan TJ, Daluiski A, et al. Online patient ratings are not correlated with total knee replacement surgeon-specific outcomes. HSS J 2018 Jul;14(2):177-180. [doi: 10.1007/s11420-017-9600-6] [Medline: 29983660]
69. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2017 Sep 8;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
70. Frost C, Mesfin A. Online reviews of orthopedic surgeons: an emerging trend. Orthopedics 2015 Apr;38(4):e257-e262. [doi: 10.3928/01477447-20150402-52] [Medline: 25901617]
71. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
72. Dorfman RG, Purnell C, Qiu C, Ellis MF, Basu CB, Kim JY. Happy and unhappy patients: a quantitative analysis of online plastic surgeon reviews for breast augmentation. Plast Reconstr Surg 2018 Dec;141(5):663e-673e. [doi: 10.1097/PRS.0000000000004268] [Medline: 29697608]
73. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
74. Lewis P, Kobayashi E, Gupta S. An online review of plastic surgeons in southern California. Ann Plast Surg 2015 May;74(Suppl 1):S66-S70. [doi: 10.1097/SAP.0000000000000517] [Medline: 25875912]

75. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
76. Donnally CJ, Roth ES, Li DJ, Maguire JA, McCormick JR, Barker GP, et al. Analysis of internet review site comments for spine surgeons: how office staff, physician likeability, and patient outcome are associated with online evaluations. Spine (Phila Pa 1976) 2018 Dec 15;43(24):1725-1730. [doi: 10.1097/BRS.0000000000002740] [Medline: 29975328]
77. Donnally CJ, McCormick JR, Li DJ, Maguire JA, Barker GP, Rush AJ, et al. How do physician demographics, training, social media usage, online presence, and wait times influence online physician review scores for spine surgeons? J Neurosurg Spine 2018 Nov 1:1-10. [doi: 10.3171/2018.8.SPINE18553] [Medline: 30497169]
78. Donnally CJ, Li DJ, Maguire JA, Roth ES, Barker GP, McCormick JR, et al. How social media, training, and demographics influence online reviews across three leading review websites for spine surgeons. Spine J 2018 Nov;18(11):2081-2090. [doi: 10.1016/j.spinee.2018.04.023] [Medline: 29709552]
79. Randhawa S, Viqar A, Strother J, Prabhu AV, Xia F, Heron D, et al. How do patients rate their radiation oncologists in the modern era: an analysis of Vitals.com. Cureus 2018 Sep 17;10(9):e3312 [FREE Full text] [doi: 10.7759/cureus.3312] [Medline: 30473945]
80. Kalagara S, Eltorai AE, DePasse JM, Daniels AH. Predictive factors of positive online patient ratings of spine surgeons. Spine J 2019 Jan;19(1):182-185. [doi: 10.1016/j.spinee.2018.07.024] [Medline: 30077043]
81. Zhang J, Omar A, Mesfin A. Online ratings of spine surgeons: analysis of 208 surgeons. Spine (Phila Pa 1976) 2018 Dec 15;43(12):E722-E726. [doi: 10.1097/BRS.0000000000002488] [Medline: 29846366]
82. Prabhu AV, Randhawa S, Clump D, Heron DE, Beriwal S. What do patients think about their radiation oncologists? An assessment of online patient reviews on Healthgrades. Cureus 2018 Feb 6;10(2):e2165 [FREE Full text] [doi: 10.7759/cureus.2165] [Medline: 29644154]
83. Jack RA, Burn MB, McCulloch PC, Liberman SR, Varner KE, Harris JD. Does experience matter? A meta-analysis of physician rating websites of orthopaedic surgeons. Musculoskelet Surg 2018 Apr;102(1):63-71. [doi: 10.1007/s12306-017-0500-1] [Medline: 28853024]
84. Daskivich T, Luu M, Noah B, Fuller G, Anger J, Spiegel B. Differences in online consumer ratings of health care providers across medical, surgical, and allied health specialties: observational study of 212,933 providers. J Med Internet Res 2018 May 9;20(5):e176 [FREE Full text] [doi: 10.2196/jmir.9160] [Medline: 29743150]
85. McGrath RJ, Priestley JL, Zhou Y, Culligan PJ. The validity of online patient ratings of physicians: analysis of physician peer reviews and patient ratings. Interact J Med Res 2018 Apr 9;7(1):e8 [FREE Full text] [doi: 10.2196/ijmr.9350] [Medline: 29631992]
86. Liu JJ, Matelski JJ, Bell CM. Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study. J Med Internet Res 2018 Mar 7;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
87. Agarwal AK, Mahoney K, Lanza AL, Klinger EV, Asch DA, Fausti N, et al. Online ratings of the patient experience: emergency departments versus urgent care centers. Ann Emerg Med 2018 Nov 1:1-8 (forthcoming). [doi: 10.1016/j.annemergmed.2018.09.029] [Medline: 30392737]
88. Geletta S. Measuring patient satisfaction with medical services using social media generated data. Int J Health Care Qual Assur 2018 Mar 12;31(2):96-105. [doi: 10.1108/IJHCQA-12-2016-0183] [Medline: 29504870]
89. Cloney M, Hopkins B, Shlobin N, Dahdaleh NS. Online ratings of neurosurgeons: an examination of web data and its implications. Neurosurgery 2018 Apr 3;83(6):1143-1152 PMID: 29618127. [doi: 10.1093/neuros/nyy064] [Medline: 29618127]
90. Farr. CNBC. As Amazon moves into health care, here's what we know—and what we suspect—about its plans URL: https://www.cnbc.com/2018/03/27/amazons-moves-into-health-what-we-know.html [accessed 2018-10-17] [WebCite Cache ID 73ErOqq8C]
91. Wald JT, Timimi FK, Kotsenas AL. Managing physicians' medical brand. Mayo Clin Proc 2017 Dec;92(4):685-686. [doi: 10.1016/j.mayocp.2017.01.005] [Medline: 28385206]
92. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. J Med Internet Res 2018 Mar 26;20(3):e99 [FREE Full text] [doi: 10.2196/jmir.8986] [Medline: 29581091]
93. Guo L, Jin B, Yao C, Yang H, Huang D, Wang F. Which doctor to trust: a recommender system for identifying the right doctors. J Med Internet Res 2016 Dec 7;18(7):e186 [FREE Full text] [doi: 10.2196/jmir.6015] [Medline: 27390219]
94. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
95. Fischer S, Emmert M. A review of scientific evidence for public perspectives on online rating websites of healthcare providers. In: In Challenges and Opportunities in Health Care Management 2015. Cham: Springer; 2014:279-290.
96. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]

Abbreviations

CMS: Center for Medicare and Medicaid Services
HCO: health care organization
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
NLP: natural language processing
POR: patient online review
PRW: patient review website

Edited by G Eysenbach; submitted 17.10.18; peer-reviewed by S Gupta, M Paul; comments to author 06.12.18; revised version received 16.12.18; accepted 31.01.19; published 08.04.19.

Please cite as:
Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL
What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews
J Med Internet Res 2019;21(4):e12521
URL: http://www.jmir.org/2019/4/e12521/
doi:10.2196/12521
PMID:

Original Paper

“But His Yelp Reviews Are Awful!”: Analysis of General Surgeons’ Yelp Reviews

Cynthia Liu[1*], BS; Meka Uffenheimer[1*], BS; Yosef Nasseri[1], MD; Jason Cohen[1], MD; Joshua Ellenhorn[1], MD

The Surgery Group of Los Angeles, Research Foundation, Los Angeles, CA, United States
[*]these authors contributed equally

Corresponding Author:
Yosef Nasseri, MD
The Surgery Group of Los Angeles
Research Foundation
Suite 880W
8635 W 3rd Street
Los Angeles, CA, 90048
United States
Phone: 1 310 289 1518
Fax: 1 310 289 1526
Email: yosefnasseri@gmail.com

Abstract

Background: Patients use Web-based platforms to review general surgeons. However, little is known about the free-form text and structured content of the reviews or how they relate to the physicians’ characteristics or their practices.

Objective: This observational study aimed to analyze the Web-based reviews of general surgeons on the west side of Los Angeles.

Methods: Demographics, practice characteristics, and Web-based presence were recorded. We evaluated frequency and types of Yelp reviews and assigned negative remarks to 5 categories. Tabulated results were evaluated using independent *t* test, one-way analysis of variance, and Pearson correlation analysis to determine associations between the number of total and negative reviews with respect to practice structure and physician characteristics.

Results: Of the 146 general surgeons, 51 (35%) had at least 1 review and 29 (20%) had at least 1 negative review. There were 806 total reviews, 679 (84.2%) positive reviews and 127 (15.8%) negative reviews. The negative reviews contained a total of 376 negative remarks, categorized into physician demeanor (124/376, 32.9%), clinical outcomes (81/376, 22%), office or staff (83/376, 22%), scheduling (44/376, 12%), and billing (44/376, 12%). Surgeons with a professional website had significantly more reviews than those without (P=.003). Surgeons in private practice had significantly more reviews (P=.002) and more negative reviews (P=.03) than surgeons who were institution employed. A strong and direct correlation was found between a surgeon’s number of reviews and number of negative reviews (P<.001).

Conclusions: As the most common category of complaints was about physician demeanor, surgeons may optimize their Web-based reputation by improving their bedside manner. A surgeon’s Web presence, private practice, and the total number of reviews are significantly associated with both positive and negative reviews.

(J Med Internet Res 2019;21(4):e11646) doi:10.2196/11646

KEYWORDS

patient satisfaction; general surgery; Los Angeles; Web-based ratings; digital health; Yelp

Introduction

One of the most ubiquitous business review websites, Yelp was established because the founder was unable to find recommendations for local physicians on the Web [1]. As customer feedback has become increasingly accessible, Web-based rating sites now have an influential impact on the impression and decisions of patients, with as many as 68% of patients turning to these resources to research or review physicians [2]. Physicians are beginning to realize that reactions and ratings detailed on these websites may impact which and how many patients visit them, as well as their overall reputation [3-5].

There are at least 33 websites where patients can describe their experience at hospitals, clinics, or clinical practices [6]. These websites range from general consumer rating websites to websites geared specifically toward the medical field. The structure of the all-purpose websites tends to afford more freedom to the commenters, whereas medical-based websites generally have a more structured format. In addition to premade surveys, several medical-based rating websites also allow reviewers to make unique remarks [6]. Yelp is one of the most used Web-based resources to review physicians [7-12].

An analysis of the content of Web-based reviews of general surgeons, including free-form content, has not been systematically described. We investigated the Yelp reviews of general surgeons in a defined region to categorize the content of the negative reviews and determine whether the number of reviews and the number of negative reviews correlated with the characteristics of the physicians. Los Angeles was chosen as the site of this study because it is home to a large variety of practices and institutions.

Methods

We identified general surgeons practicing on the west side of Greater Los Angeles using The Medical Board of California Web-based database [13] and InfoUSA (Papillion, Nebraska), a marketing company that provides contact databases and mailing lists. The physicians practicing in the 31 zip codes on the west side of Los Angeles were examined. The active practice status of the physician was determined using The Medical Board of California's License verification Web-based tool [13]. Those surgeons who were still in training, retired, and those without an active medical license were eliminated. Physician gender, years since graduating from medical school, and medical school attended were identified. A Google search was performed to determine if the physician had a medical practice website.

A physician was considered to have a Yelp page if they could be found and reviewed on a page designated for an institution, clinical practice, or the physician. The following was documented: presence of a Yelp page, number of reviews, number of positive reviews, and number of negative reviews. Yelp users rate an institution or physician on a 5-star system, with 5 stars defined as "Woohoo! As good as it gets!" and 1 star defined as "Eek! Methinks not." Yelp defines 4 stars as "Yay! I'm a fan." As 4 stars was less than a perfect star rating, a negative review was defined as a review that contained at least 1 negative remark and had a rating of 4 stars or fewer. Yelp defines 3 stars as A-OK, and because it is not a 4- or 5-star review, a level of mediocrity is implied [14]. A positive review was defined as any 5-star or 4-star review that contained no negative remarks.

Negative reviews were further characterized. Each commenter's username, date of review, and star rating was noted. Free-form reviews were manually tabulated, categorized, and resolved by 2 independent reviewers and were empirically divided into 5 categories modified from previously defined categories [15]: Scheduling (doctor availability and punctuality), billing, office and staff (staff friendliness or professionalism, staff presence, and office décor or location), clinical outcome (correct diagnosis or treatment, technical skill, treatment of unforeseen complications, if additional treatment was needed, and follow-up care), and physician demeanor (education, empathy, bedside manner, professionalism, preparedness or organization, time with doctor, communication skills, shared decision making, and general impression). Remarks within negative reviews were noted to be negative, neutral, or positive.

Once the raw data were collected, summary statistics were calculated using univariable analyses. An independent *t* test, which determines whether there is a statistically significant difference between the means in 2 unrelated groups, was run to determine whether or not certain factors (possession of professional website, private practice or institution employed, gender, and medical school outside of the United States) impacted the number of total reviews or negative reviews of a surgeon. One-way analysis of variance analysis, which determines whether there is a statistically significant difference between the means in more than 2 unrelated groups, was used to determine if zip code or the number of years since graduating from medical school impacted a surgeon's number of total reviews or negative reviews. A Pearson correlation, which measures the linear correlation between 2 variables, was performed to determine if there was any relationship between a give surgeon's total number of reviews and quantity of negative reviews. Statistical Package for the Social Sciences statistical software was used.

Results

We identified 146 practicing general surgeons on the west side of Greater Los Angeles. A total of 33 (22.6%, 33/146) surgeons were female and 113 (77.4%, 113/146) were male. Moreover, 55 (37.7%, 55/146) surgeons were in private practice and 91 (69.2%, 91/146) were institution employed. In addition, 19 (13.0%, 19/146) surgeons went to medical school outside of the United States. Furthermore, 42 (28.8%, 42/146) physicians had a professional website (Table 1).

A total of 59 (40.4%, 59/146) surgeons had a Yelp page. Moreover, 51 physicians (34.9% of all surgeons and 86% of those with a Yelp) had at least 1 review. Of the 59 physicians who had a Yelp page, 29 (49%, 29/59) had at least 1 negative review and 48 (81%, 48/59) had at least 1 positive review (Figure 1).

There was a total of 806 documented reviews. Of these reviews, 679 (84.2%) were positive and 126 (15.6%) were negative. Within the 126 negative reviews, there were 376 negative remarks: 124 (32.9%) concerning physician demeanor, 81 (21.5%) on clinical outcomes, 83 (22.0%) regarding the office or staff, 44 (11.7%) about scheduling, and 44 (11.7%) in relation to billing (Table 1).

Table 1. Demographics (N=146).

Characteristics	Statistics, n (%)
Male	113 (77.4)
Private practice	55 (38)
Institutional employed	91 (69.2)
Physicians with website	42 (28.7)
Foreign medical school graduate	19 (13.0)
Physicians with a Yelp page	59 (40.4)
Physicians reviewed on Yelp	51 (34.9)
Physicians with negative reviews	29 (19.8)

Figure 1. Distribution of reviews. The chart on the left shows the breakdown of positive reviews among the 59 physicians who had a Yelp page, whereas the chart on the right shows the breakdown of negative reviews. The number of negative and positive reviews is shown along with the number of physicians who had that number of negative and positive reviews, respectively.

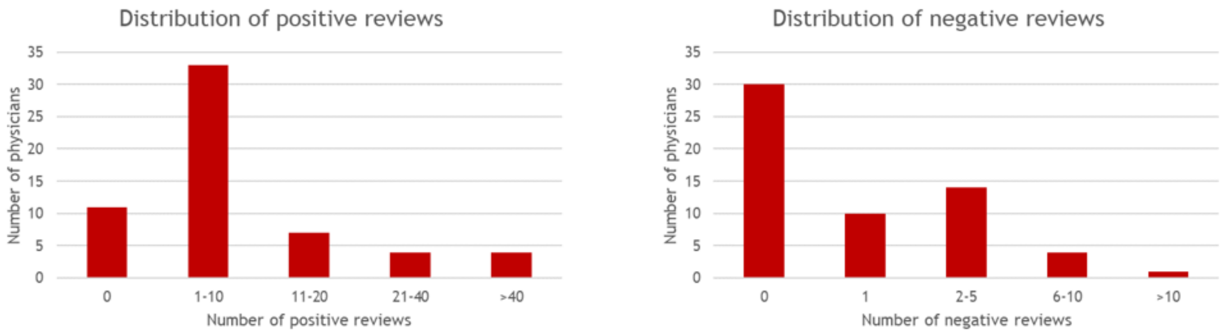


Table 2. Statistical analysis via 2-way Pearson correlation analysis.

Reviews	Number of reviews, mean (SD)	*P* value
Total reviews (n=806)	5.6 (18.08)	.001
Negative reviews (n=126)	0.86 (4.33)	.001

The existence of a professional website as well as the type of employment had a significant impact on the total number of reviews (Tables 2-4). Those with a professional website had significantly more overall reviews compared with those without (mean 16.36, SD 31.0 vs mean 1.14, SD 3.23; *P*=.003). Two-thirds of physicians in private practice had reviews, compared with one-third of physicians employed by an institution (mean 18.0, SD 31.4 vs mean 0.96, SD 3.18; *P*=.002). Being in private practice was also significantly associated with the number of negative reviews (mean 3.11, SD 8.13 vs mean 0.09, SD 0.35; *P*=.03). As such, private practice was significantly associated with both more overall reviews and more negative reviews. Furthermore, there was a trend toward a correlation between a greater number of reviews and a greater number of years since graduation from medical school (*P*=.05).

A 2-tailed Pearson correlation was performed to evaluate the association between the total number of reviews and the number of negative reviews. The correlation between the total number of reviews and negative reviews was strong (*r*=.862; *P*<.001).

The total number of reviews was not impacted by a surgeon's gender (male: mean 4.2, SD 11.6 vs female: mean 9.9, SD 31.3; *P*=.32), whether they attended medical school outside the United States (non-United States: mean 1.6, SD 3.20 vs United States: mean 6.1, SD 19.3; *P*=.32), or the practice zip code (*P*=.20).

The likelihood of having negative reviews did not significantly differ between surgeons with a professional website versus surgeons without a professional website (mean 2.3, SD 7.80 vs mean 0.28, SD 0.98; *P*=.10). Similarly, the number of negative reviews was not significantly affected by a surgeon's gender (male: mean 0.48, SD 1.22 vs female: mean 2.18, SD 8.79; *P*=.28), whether or not they attended medical school outside the United States (non-United States: mean 0.6, SD 1.54 vs United States: mean 0.91, SD 4.6; *P*=.76), the zip code they practiced in (*P*=.77), or the number of years since the surgeon graduated from medical school (*P*=.85).

Table 3. Statistical analysis via independent *t* test.

Variable	Number of physicians with any reviews	Reviews			Negative reviews		
		Total, N	Mean (SD)	*P* value	Total, N	Mean (SD)	*P* value
Private practice physicians (n=55)	34	720	18 (31.4)	.002	118	3.11 (8.13)	.03
Institution employed physicians (n=91)	17	86	0.96 (3.18)	.002	8	0.09 (0.35)	.03
Medical school out of United States (n=19)	5	31	1.6 (3.20)	.32	11	0.6 (1.54)	.76
Medical school in United States (n=127)	46	775	6.1 (19.3)	.32	115	0.91 (4.60)	.76
Male (n=113)	38	480	4.2 (11.6)	.32	54	0.48 (1.22)	.28
Female (n=33)	13	326	9.9 (31.3)	.32	72	2.18 (8.79)	.28
Possess professional website (n=42)	28	687	16 (31.0)	.003[a]	97	2.3 (7.80)	.10
Do not possess professional website (n=104)	23	119	1.1 (3.23)	.003	29	0.28 (0.98)	.10

Table 4. Statistical analysis via 1-way analysis of variance analysis.

Variable	*P* value	
	Comparing total number of reviews	Comparing number of negative reviews
Number of years since graduating from medical school	.05	.85
Zip code	.20	.7

Discussion

Our finding that the majority of Web-based reviews were positive supports prior findings [10,16]. Unfortunately, negative reviews do exist and can have a damaging effect on a physician's reputation and practice [3,5,9,17,18]. The most common category of Yelp complaints was physician demeanor. As such, surgeons may optimize their Web-based reputation by improving their bedside manner. A surgeon's type of practice (ie, private practice or institution employed) and the total number of reviews were significantly associated with more negative reviews. A surgeon's type of practice and the possession of a personal website were significantly associated with more total Yelp reviews.

Of the negative reviews, the most common category was criticism of physician demeanor. Patient perception of physician demeanor is a factor about which physicians may have some control. Thus, it might be possible to improve one's reviews by enhancing patient-physician interactions. This includes being better prepared for each consultation, spending sufficient time with patients, clearly communicating the plan of care and disease processes affecting each patient, and displaying empathy. Over one-fifth of the complaints centered around clinical outcomes and another one-fifth concentrated on the office or staff. Although physicians are not always in control of clinical outcomes, keeping one's skills up to date and practicing evidence-based medicine may improve clinical outcomes. To reduce the number of negative Web-based reviews, it is probably important to treat unforeseen complications quickly and empathically. In addition, hiring courteous office staff and optimizing the aesthetic appearance of one's practice environment may enhance the quality of Web-based reviews.

Our study found that Web presence significantly impacted a physician's total number of reviews. Presence of a website, personal or private practice, was shown to have a significant effect on the total number of reviews a physician received. The mean number of reviews for those with a professional website was significantly greater than the number of reviews for surgeons without a website, indicating that an increased Web presence leads to more Web-based reviews. General surgeons in private practice had significantly more reviews overall (*P*=.002) and significantly more negative reviews (*P*=.03) than those who were institution employed. In addition, the total number of reviews was strongly correlated with the number of negative reviews. This suggests that although a Web-based presence may be important in enhancing a surgeon's reputation, it may also be detrimental, depending on the content of the individual reviews.

There are several limitations to this study. Restricting physicians examined to a single geographic area is descriptive for the region but is also limiting in size and scope. Due to the presence of large health institutions in the west side of Los Angeles, the number of surgeons with reviews may be artificially low because institution-based physicians may have a much more limited Web presence and thus fewer Web-based reviews. Yelp was the only Web-based rating website examined, excluding any reviews and Web-based presence surgeons may have had on other rating websites [7]. There is inherent response bias as patients make the active decision whether to write a review or not. This study only aimed to examine patient perception of physicians as well as any perceived problems they encountered in their experiences, and as such, we cannot comment on how this reflects on the quality of the physicians themselves [17,19].

Future studies are needed to determine if the trends and correlations found are applicable to surgeons practicing in other and larger geographic regions. Other specialties need to be examined to see if there is a difference in the number and type of reviews and Web presence among other specialties. Most importantly, the impact of Yelp reviews on a surgeon's practice needs to be assessed.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

References

1. Yelp Official Blog: Yelp, Inc; 2013. Study: Yelp ratings work for hospitals, too URL: https://www.yelpblog.com/2013/03/study-yelp-ratings-work-for-hospitals-too [accessed 2019-03-17] [WebCite Cache ID 76w1locCq]
2. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and analyzing patient experiences of health care from social media. JMIR Res Protoc 2015;4(3):e78 [FREE Full text] [doi: 10.2196/resprot.3433] [Medline: 26137885]
3. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
4. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
5. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. J Gen Intern Med 2017 Jun;32(6):626-631. [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]
6. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
7. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. J Am Med Assoc 2017 Dec 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
8. Geletta S. Measuring patient satisfaction with medical services using social media generated data. Int J Health Care Qual Assur 2018 Mar 12;31(2):96-105. [doi: 10.1108/IJHCQA-12-2016-0183] [Medline: 29504870]
9. Dorfman RG, Purnell C, Qiu C, Ellis MF, Basu CB, Kim JYS. Happy and unhappy patients: a quantitative analysis of online plastic surgeon reviews for breast augmentation. Plast Reconstr Surg 2018 Dec;141(5):663e-673e. [doi: 10.1097/PRS.0000000000004268] [Medline: 29697608]
10. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
11. Ramkumar PN, Navarro SM, Chughtai M, La T, Fisch E, Mont MA. The patient experience: an analysis of orthopedic surgeon quality on physician-rating sites. J Arthroplasty 2017 Dec;32(9):2905-2910. [doi: 10.1016/j.arth.2017.03.053] [Medline: 28455178]
12. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
13. Medical Board of California. 2018. Medical Board of California Breeze License Verification URL: http://www.mbc.ca.gov/Breeze/License_Verification.aspx [accessed 2019-03-14] [WebCite Cache ID 76sDdjArX]
14. Woolf M. minimaxir. The Statistical Difference Between 1-Star5-Star Reviews on Yelp URL: https://minimaxir.com/2014/09/one-star-five-stars/ [accessed 2019-03-14] [WebCite Cache ID 76sDipxEh]
15. Asanad K, Parameshwar PS, Houman J, Spiegel BM, Daskivich TJ, Anger JT. Online physician reviews in female pelvic medicine and reconstructive surgery: what do patients really want? Female Pelvic Med Reconstr Surg 2018;24(2):109-114. [doi: 10.1097/SPV.0000000000000503] [Medline: 29474282]
16. Cloney M, Hopkins B, Shlobin N, Dahdaleh NS. Online ratings of neurosurgeons: an examination of web data and its implications. Neurosurgery 2018 Apr 03;83(6):1143-1152. [doi: 10.1093/neuros/nyy064] [Medline: 29618127]
17. Chen J, Presson A, Zhang C, Ray D, Finlayson S, Glasgow R. Online physician review websites poorly correlate to a validated metric of patient satisfaction. J Surg Res 2018 Jul;227:1-6. [doi: 10.1016/j.jss.2018.01.037] [Medline: 29804840]
18. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]

19. Haskins IN, Krpata DM, Rosen MJ, Perez AJ, Tastaldi L, Butler RS, et al. Online surgeon ratings and outcomes in hernia surgery: an Americas Hernia Society quality collaborative analysis. J Am Coll Surg 2017 Nov;225(5):582-589. [doi: 10.1016/j.jamcollsurg.2017.08.007] [Medline: 28838868]

Edited by G Eysenbach; submitted 23.07.18; peer-reviewed by T Daskivich, S Zheng; comments to author 04.10.18; revised version received 11.11.18; accepted 23.01.19; published 30.04.19.

Please cite as:
Liu C, Uffenheimer M, Nasseri Y, Cohen J, Ellenhorn J
"But His Yelp Reviews Are Awful!": Analysis of General Surgeons' Yelp Reviews
J Med Internet Res 2019;21(4):e11646
URL: http://www.jmir.org/2019/4/e11646/
doi:10.2196/11646
PMID:31038463

Original Paper

Developing a Physical Activity Ontology to Support the Interoperability of Physical Activity Data

Hyeoneui Kim[1], PhD, MPH, RN, FAAN; Jessica Mentzer[1], BSN, RN; Ricky Taira[2], PhD

[1]School of Nursing, Duke University, Durham, NC, United States

[2]Department of Radiological Science, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Hyeoneui Kim, PhD, MPH, RN, FAAN
School of Nursing
Duke University
307 Trent Drive
Durham, NC, 27710
United States
Phone: 1 919 684 7534
Email: hyeoneui.kim@duke.edu

Abstract

Background: Physical activity data provides important information on disease onset, progression, and treatment outcomes. Although analyzing physical activity data in conjunction with other clinical and microbiological data will lead to new insights crucial for improving human health, it has been hampered partly because of the large variations in the way the data are collected and presented.

Objective: The aim of this study was to develop a Physical Activity Ontology (PACO) to support structuring and standardizing heterogeneous descriptions of physical activities.

Methods: We prepared a corpus of 1140 unique sentences collected from various physical activity questionnaires and scales as well as existing standardized terminologies and ontologies. We extracted concepts relevant to physical activity from the corpus using a natural language processing toolkit called Multipurpose Text Processing Tool. The target concepts were formalized into an ontology using Protégé (version 4). Evaluation of PACO was performed to ensure logical and structural consistency as well as adherence to the best practice principles of building an ontology. A use case application of PACO was demonstrated by structuring and standardizing 36 exercise habit statements and then automatically classifying them to a defined class of either sufficiently active or insufficiently active using FaCT++, an ontology reasoner available in Protégé.

Results: PACO was constructed using 268 unique concepts extracted from the questionnaires and assessment scales. PACO contains 225 classes including 9 defined classes, 20 object properties, 1 data property, and 23 instances (excluding 36 exercise statements). The maximum depth of classes is 4, and the maximum number of siblings is 38. The evaluations with ontology auditing tools confirmed that PACO is structurally and logically consistent and satisfies the majority of the best practice rules of ontology authoring. We showed in a small sample of 36 exercise habit statements that we could formally represent them using PACO concepts and object properties. The formal representation was used to infer a patient activity status category of sufficiently active or insufficiently active using the FaCT++ reasoner.

Conclusions: As a first step toward standardizing and structuring heterogeneous descriptions of physical activities for integrative data analyses, PACO was constructed based on the concepts collected from physical activity questionnaires and assessment scales. PACO was evaluated to be structurally consistent and compliant to ontology authoring principles. PACO was also demonstrated to be potentially useful in standardizing heterogeneous physical activity descriptions and classifying them into clinically meaningful categories that reflect adequacy of exercise.

(J Med Internet Res 2019;21(4):e12776) doi:10.2196/12776

KEYWORDS

exercise; leisure activities; health information interoperability; terminology as topic

Introduction

Challenges in Reusing Physical Activity Data

Undoubtedly, a healthy lifestyle, especially being physically active, is paramount to healthy living. Numerous scientific studies have shown the direct impact of physical activity on disease onset and progress as well as treatment outcomes [1-7]. Although analyzing physical activity data in conjunction with other clinical and microbiological data will lead to new insights crucial for improving human health, its execution is challenging because of the large variation in the way the data are collected and presented.

The first challenge relates to the heterogeneous nature of the acquired data for measuring and assessing physical activity. High-resolution temporal samples of physical activity data captured through personal sensor devices are now becoming increasingly available and feasible for ubiquitous monitoring. Questionnaire-based descriptive measures are also widely used to assess one’s overall exercise habits and factors affecting one’s ability and desire to be physically active. These questionnaire-based measures complement objective measures associated with sensor devices. Each measure poses challenges to reusing the data that it generates. This study concerns improving the reusability of the descriptive measures generated based on questionnaires and assessment scales.

The second challenge relates to the usability of the data. Physical activity data are not free from a common barrier to utilizing text data, which is transforming the data into a computable—that is, structured and standardized—format [8,9]. For example, physical activity data are usually described with nontechnical terms in a lengthy sentence and often buried in narrative notes produced during a clinical encounter. A sample clinical question asked might appear as the following [10]:

> *Over the past 7 days, how often did you engage in light sport or recreational activities such as bowling, golf with a cart, shuffleboard, fishing from a boat or pier or other similar activities?*

Use of lengthy descriptive sentences can help to clearly convey the intention of the question by minimizing the room for misinterpretation. However, it also creates a challenge to systematically analyzing such data in conjunction with other clinical and biological data.

Limitations of the Existing Standardization Approaches to Representing Physical Activity Data

Common data elements (CDEs) are used in several disciplines for standardizing data collected with assessment scales and questionnaires and have been widely adopted for acquiring self-reported data including physical activity information [11]. The consensus measures for Phenotypes and eXposures (PhenX) Toolkit is a collection of standard measurement protocols that can be used in biomedical research, developed through the cooperative agreement between RTI international and National Institute of Health [12]. PhenX offers a number of standardized scales and questionnaires recommended for collecting physical activity data [12]. Part of the PhenX measures are now included in Logical Observation Identifiers Names and Codes [13]. The CDE repository of the National Library of Medicine allows users to search standardized data elements in the biomedical domain and provides rich metadata on the queried data element, including standardized concept codes [14].

CDEs are an effective method of standardizing questionnaire-based data. However, there are a number of studies involving locally developed questionnaires not covered by the CDE-based standardization efforts. Often, there is a substantial informational overlap within the various questionnaires. For example, there are questions that relate to highly similar topics but belong to different questionnaires and are thus treated as distinctive CDEs. For example, “In the past 7 days, how many days were you physically active for 10 minutes or more?” and “Physical activity 20 minutes per day during week count” both ask for the number of days in a week that a person is physically active although different activity durations are indicated. The former is a question from the Quality of Life in Neurological Disorders questionnaire [15], and the latter is from the National Institute of Neurological Disorders and Stroke questionnaire [16]. Systematically recognizing their similarities will facilitate interoperability of physical activity–related data.

Physical activity is 1 of the 9 social and behavioral health domains that need to be incorporated into electronic health records (EHRs) in a structured format, as recognized by the Office of National Coordinators (ONC) and the Institute of Medicine (IOM—currently the National Academy of Medicine) [17]. Furthermore, ONC and IOM recognized the following 2 salient questions from Exercise Vital Sign [18] as candidate measures to assess physical activity:

1. On average, how many days per week do you engage in moderate to strenuous exercise (such as walking fast, running, jogging, dancing, swimming, biking, or other activities that cause a light or heavy sweat)?
2. On average, how many minutes do you engage in exercise at this level?

These 2 questions certainly provide minimum necessary information on a patient’s overall exercise habit. However, incorporating physical activity as a care regimen or investigating how it affects health outcomes requires a more detailed and diverse representation of one’s physical activity level. In addition, the challenges to reusing the existing physical activity data described for a patient within the EHR remain formidable, including issues related to ambiguity and semantic inconsistency. Furthermore, these data are still largely buried in clinical narrative texts with highly variable forms of expression. Thus, given the increasing clinical awareness of the importance of physical activity assessment, there is a pressing need to explore defining an expanded representation for physical activity data that complements and consolidates existing standardization efforts.

Gaps in the Existing Ontologies for Physical Activity Data

Many ontologies and standardized terminologies cover aspects of the physical activity domain, but concept coverage remains incomplete. We reviewed existing relevant ontologies and terminology systems to benchmark their structure and to aggregate relevant concepts for our proposed Physical Activity Ontology (PACO). The Semantic Mining of Activity, Social, and Health data (SMASH) ontology contains 74 concept classes that cover the concepts related to social activities and network [19]. SMASH has a well-developed Physical Activity type hierarchy that is divided into Athletic Sports, Exercise, and Occupational Activity. However, its concept coverage on activity types is quite limited, and it does not offer modifier concepts required to describe intensity and amount of physical activity. The Ontology for assessing Physical Activity and Sedentary Behavior (OPA) provides formal expressions for the various domains of concepts relevant to physical activity [20]. OPA focuses on formally representing main top-level concept classes such as *TemporalEntity*, *SpaceEntity*, *Person*, *SocialContext*, etc, and concept properties that link the concept classes. OPA is designed to be used in conjunction with other terminology systems. Therefore, it does not include detailed concepts that belong to the classes. The Ontology of Physical Exercises (OPEs) is a Web Ontology Language (OWL) formatted ontology developed to support consistent representation of exergame data [21]. OPE has comprehensive coverage of concepts important to represent exergame data including game equipment types, health outcomes, engaged musculoskeletal systems, and disease and injuries. OPE also contains some exercise concepts as general categories such as aerobic exercise, isometric exercise, light exercise, etc. Naturally, OPE has many limitations to be considered as an ontology to support representing nongame-based physical activities with sufficient details.

Study Aim

Consistent and unambiguous representation of physical activity data is essential to draw increased insights that support patient care and health outcome research. There is a need to identify a robust and systematic approach to structuring and standardizing heterogeneous descriptions on one's physical activity expressed with various measures. The aim of this study was to develop an ontology for physical activity with the concepts important to describe clinically meaningful characteristics of people's physical activity.

Methods

Data Sources

We collected 1140 unique questions and sentences on physical activity from 92 questionnaires and forms such as Healthy Living Questionnaire [22], Rapid Assessment of Physical Activity [23], Two Question Physical Activity Assessment [24], etc. The full list of questions and sentences analyzed in this study is provided in Multimedia Appendix 1.

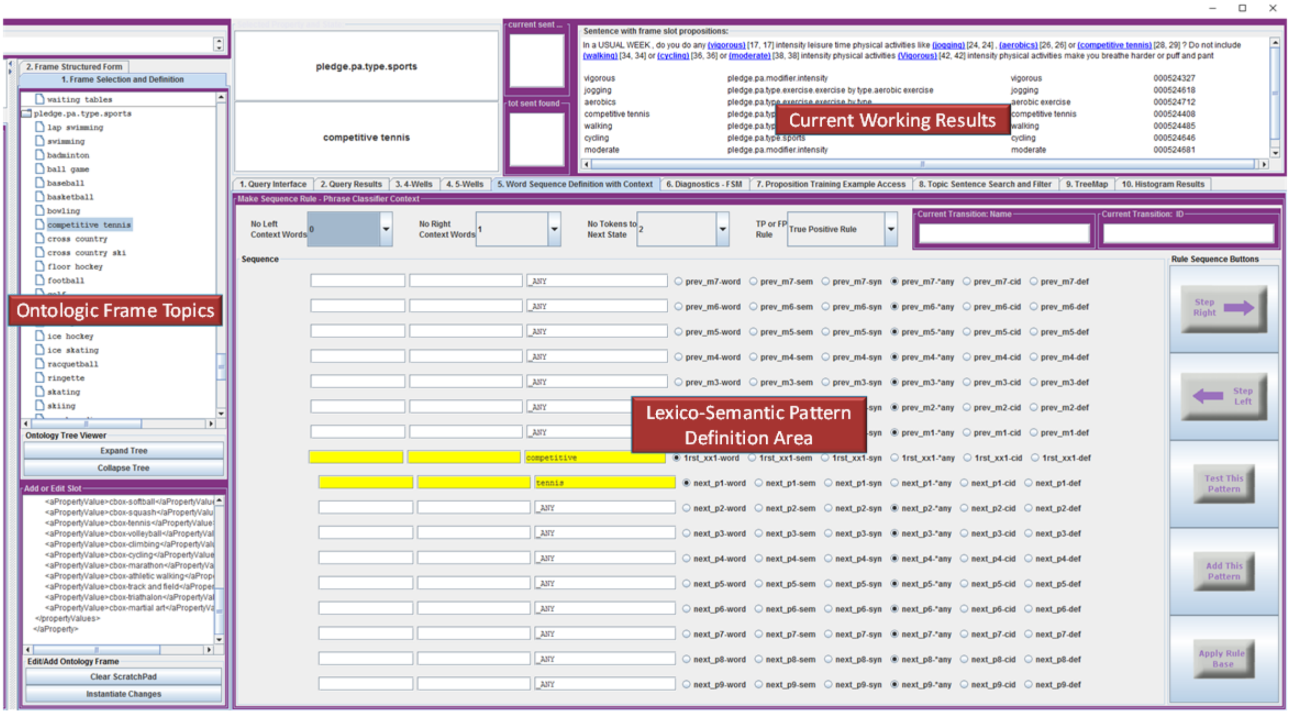
Multipurpose Text Processing Tool

Exploration of the concepts and terms referenced within the 1140 unique questions were facilitated using a natural language processing (NLP) tool called Multipurpose Text Processing Tool (MUTT) developed by the Medical Imaging Informatics Group at the University of California Los Angeles. This NLP environment is designed to allow developers to both define the ontologic elements and structure of the target domain and link corresponding NLP lexico-syntactic-semantic patterns to identify them in free text. This approach is similar to recent ontology-driven NLP applications such as is employed in the OpenDMAP project [25]. The NLP pattern acquisition aspect is data driven, similar in this respect to the knowledge discovery methods reported in another study [26]. Details of the core NLP system can be found in other studies [27-29].

The main MUTT module interface is shown in Figure 1. The first step is to define a topic class (eg, PACO) using a free text XML editor. This high-level class allows various NLP extraction modules related to this class to be activated. The next step involves defining the hierarchy of concepts under the topic class. Once a user has defined an ontological class definition (eg, exercise.equipment) and its possible instances (eg, treadmill, rowing machine, or elliptical), the next step involves defining NLP extraction patterns to identify such instances in free text. This step allows the system to precompile a knowledge source that serves as a mapping between ontological instances and all their associated lexical variant patterns.

Practically, users can instantiate extraction patterns best by viewing training examples (ie, a particular question) and using the Figure 1 interface to define concept detection patterns. With this approach, the following steps are performed: (1) user selects a training sentence; (2) the system tokenizes the sentence and displays all ontologic concepts it currently can extract (see the Current Working Results area of Figure 1); (3) the user examines the results and can decide if there are any ontologic terms missing from the extracted sentence results; (4) as part of the results for the selected sentence, the system presenting the user with a scrollable table listing 1 word token per line with column fields corresponding to exclusively selectable matching attributes of the token including its exact surface string, semantic class, part-of-speech class, wildcard and the user definable and selectable attributes of predefined morphological features, and token-level regular expression; and (5) the user can define an extraction pattern across multiple tokens (ie, multiple word phrases) in this manner, with each token within the sequence specified separately. The interface also allows users to specify left and right context tokens to partially address issues of semantic ambiguity within the context of the question. Patterns can be associated with either a true positive or a false positive match. Many such patterns can be associated with a single ontologic concept.

Figure 1. Screenshot of the multipurpose text processing tool user-interface environment. The Ontologic Frame Topics box displays an ontology specification with XML and tree representations; Users specify lexico-syntactic-semantic patterns of the text corresponding to ontologic property instance in Lexico-Semantic Pattern Definition Area; The Current Working Results box provides immediate user feedback to specified patterns on training sentence. Not shown are user's screens to view training sentences, training status, and automated pattern discovery results.



Sentence with frame slot propositions:

In a USUAL WEEK , do you do any (vigorous) [17, 17] intensity leisure time physical activities like (jogging) [24, 24] , (aerobics) [26, 26] or (competitive tennis) [28, 29] ? Do not include (walking) [34, 34] or (cycling) [36, 36] or (moderate) [38, 38] intensity physical activities (Vigorous) [42, 42] intensity physical activities make you breathe harder or puff and pant

```
vigorous            pledge.pa.modifier.intensity                                vigorous            000524327
jogging             pledge.pa.type.exercise.exercise by type.aerobic exercise   jogging             000524618
aerobics            pledge.pa.type.exercise.exercise by type                    aerobic exercise    000524712
competitive tennis  pledge.pa.typ                                               competitive tennis  000524408
walking             pledge.pa.typ                                               walking             000524485
cycling             pledge.pa.type.sports                                       cycling             000524646
moderate            pledge.pa.modifier.intensity                                moderate            000524681
```

Table 1. Sequence of training and testing steps for term extraction and the number of sentences used for training and testing.

Step	Text data	Annotator	Task
1	Set 1 (n[a]=100)	Human	Developing baseline semantic models
2	Set 2 (n=400)	Human	Round 1 training
3	Set 3 (n=100)	MUTT[b]	Round 1 testing
4	Test number 1 results	Human	Reviewing and analyzing the round 1 testing results
5	Set 4 (n=300)	Human	Round 2 training
6	Set 5 (n=240)	MUTT	Round 2 testing
7	Test number 2 results	Human	Reviewing and analyzing the round 2 testing results

[a] n=number of sentences included in the annotation set.

[b] MUTT: multipurpose text processing tool.

Harvesting Terms

We first structured 100 questions by annotating key concepts based on a preliminary concept model from a previous effort [30]. This preliminary model comprises 3 concept classes including activity type, modifiers, and facilitating and inhibiting factors. The model also includes semantic relations among these classes. Through this initial analysis, we further specified and expanded the base concept classes in the preliminary concept model by creating multiple child classes. Activity types were divided into *exercise*, *daily activity (including household chores)*, and *leisure/recreational activities*. Modifiers were further detailed into *amount*, *frequency*, and *intensity*. We also identified additional concept classes important to capture the concepts describing people's activity level during this initial annotation. For example, *exercise location*, *exercise equipment*, and *fitness program/classes* were added. Note that our goal of modeling at this point was to be as comprehensive as possible.

The baseline model was populated as the initial semantic frame model in MUTT. The remaining 1040 sentences were annotated using MUTT. Text annotation and the semantic model augmentation process were performed iteratively by 2 annotators as illustrated in Table 1.

Building an Ontology

The ontologic frame definitions manually specified using the MUTT interface were converted from XML to OWL and subsequently imported as a baseline ontology into Protégé, an ontology authoring tool developed by Stanford Center for Biomedical Informatics Research at the University [31]. We also searched the National Center for Biomedical Ontology BioPortal [32] for any ontology relevant to physical activity to incorporate additional concepts and relations. BioPortal is a Web service that allows users to upload, search, and access biomedical ontologies, developed and maintained at the Stanford University. We reviewed ontologies retrieved with the search term *exercise* and *physical activity*, which yielded 48 ontologies or standardized terminology systems including Systematized Nomenclature of Medicine-Clinical Terms [33], Medical Subject Headings [34], the National Cancer Institute Thesaurus [35], and the Read Clinical Terminology Version 2 [36]. Most of these ontologies have a small substructure (a branch or a single class) relevant to physical activity. We found that majority of the concepts found in these ontologies were already included in our baseline ontology. Through this activity, we added 1 new concept, *high intensity interval training*. Cross-referencing previous ontologies in addition provided an opportunity to ensure that our concepts were appropriately phrased and placed at an appropriate level in the hierarchy.

Naming conventions of class labels used the singular form of nouns and verbs with the first letter of a word being capitalized. An underscore is inserted between words for a multiword label (eg, *Ice _ hockey* and *Circuit_training*). Sport or exercise names are included in a noun form. We adopted a gerund form if a concept has only a verb form. Instances (Individuals in Protégé) are labeled in all lowercase and likewise, an underscore is inserted between words for a multiword instance label (eg, *high_impact* and *make_you_puff_and_pant*). Property names followed a camel case style (eg, *hasIntensity* and *hasActivityEffect*).

Overall, 2 general activity classes were defined at a top level that differentiated daily (*Daily_living_activity)* versus leisure (*Exercise_leisure_activity)* activities. These classes were then further divided into a number of subclasses. There are multiple ways to categorize the activity names under *Exercise_leisure_activity*. For example, dancing is a physically active leisure activity and at the same time can be considered a comprehensive exercise that helps with endurance, flexibility, balance, and bone and muscle strength. *Cross_country_ski* can be classified as a winter outdoor sport in addition to the categories defined by its exercise effects.

To efficiently handle the complexities in classifying activity types, we adopted an asserted hierarchy and a defined hierarchy. We asserted the hierarchy of *Exercise_leisure_activity* using apparent and more generic subsumptive relations. For example, the *Ballgame* class contains different ball games such as *Soccer*, *Baseball*, *Tennis*, etc. Similarly, the *Running* class contains various exercise and/or sports characterized by running such as *Jogging*, *Treadmill_running*, *Sprinting*, *Marathon*, etc. Additional activity type classes were created as a defined hierarchy to incorporate a few common ways of categorizing exercise and leisure activities such as by exercise effects, by indoor or outdoor activity, winter activity, and water activity. These multiple views of organizing the ontology were implemented using a multiple-inheritance structure. We defined each named activity under the *Exercise_leisure_activity* class with the 3 properties of *hasActivityEffect*, *hasActivityLocation*, and *hasActivityRequiredCondition*. We then generated an inferred hierarchy where these named activities are classified under the defined classes. This inferred hierarchy was specified using FaCT++, a Web Ontology Language Description Logic (OWL-DL) reasoner available in Protégé [37]. FaCT++ is an open-source software developed by Dmitry Tsarkov and Ian Horrocks at the University of Manchester [37]

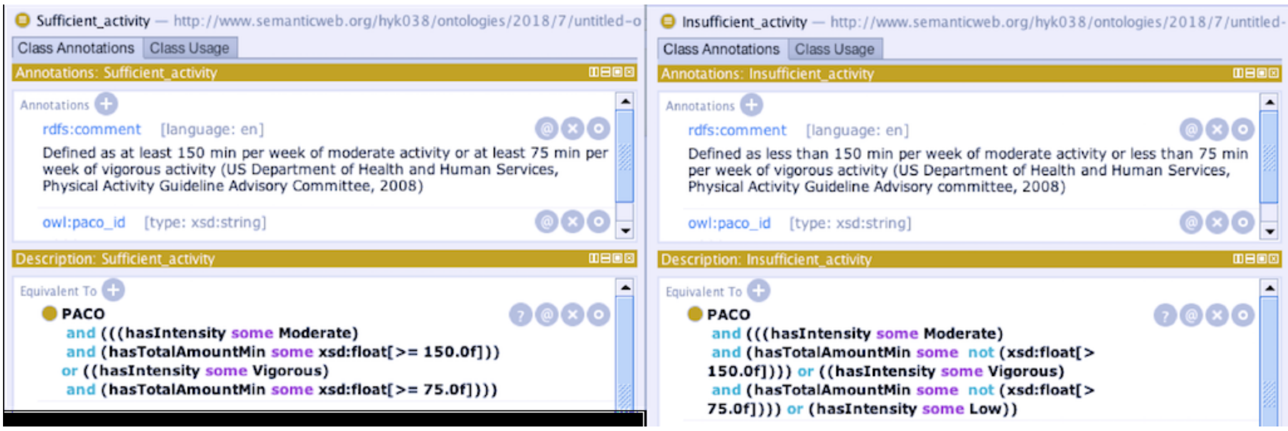
Ontology Evaluation

We checked the logical and structural quality of PACO first using the Ontology Debugger plugin available in Protégé [38]. We also tested PACO with the Ontology Pitfall Scanner! (OOPS!) tool [39] to ensure its compliance to the ontology authoring principles in addition to the structural quality. OOPS! is a Web-based tool from the Ontology Engineering Group of the Technical University of Madrid that examines an ontology against 33 common pitfalls as compared with state-of-the-art principles for ontology construction. These pitfalls cover not only logical and structural issues but also usability and documentation concerns [40].

As an additional quality assurance effort, we tested the system's ability to identify ontologic concepts from various free-text physical activity descriptions. Clinically, 2 important outcome nodes were added for this step that represent adequacy of exercise, that is, *Sufficient_exercise* and *Insufficient_exercise*. These classes were defined using the following 2 properties: *hasIntensity* that captures an intensity level and *hasTotalAmountInMin* that captures total weekly exercise amount measured in minutes (see Figure 2). The definitions of the 2 exercise levels were developed based on the physical activity guideline provided by the US Department of Health and Human Service in 2008 [41].

We collected 36 descriptions on usual exercise habits from Web consumer-oriented behavioral health articles (eg, World Health Organization and American Heart Association) and from a convenience sample of 30 people (ie, friend, family, and colleagues of the authors). We formally defined these 36 statements using *hasIntensity* and *hasTotalAmountInMin*, and then added them to PACO as an instance (ie, individual in Protégé) of a specific exercise type. We evaluated the concept coverage of the *Exercise_leisure_activity* class by identifying an exercise type that an instance belongs to and by populating the intensity property with an intensity concept from the Intensity class. We classified the 36 exercise statements into 1 of the 2 exercise level classes using FaCT++.

Figure 2. Defining exercise level classes.



rdfs:comment [language: en]

Defined as at least 150 min per week of moderate activity or at least 75 min per week of vigorous activity (US Department of Health and Human Services, Physical Activity Guideline Advisory Committee, 2008)

owl:paco_id [type: xsd:string]

```
PACO
 and (((hasIntensity some Moderate)
 and (hasTotalAmountMin some xsd:float[>= 150.0f]))
or ((hasIntensity some Vigorous)
 and (hasTotalAmountMin some xsd:float[>= 75.0f])))
```

rdfs:comment [language: en]

Defined as less than 150 min per week of moderate activity or less than 75 min per week of vigorous activity (US Department of Health and Human Services, Physical Activity Guideline Advisory committee, 2008)

owl:paco_id [type: xsd:string]

```
PACO
 and (((hasIntensity some Moderate)
 and (hasTotalAmountMin some  not (xsd:float[>
150.0f]))) or ((hasIntensity some Vigorous)
 and (hasTotalAmountMin some  not (xsd:float[>
75.0f]))) or (hasIntensity some Low))
```

Results

Natural Language Processing Term Extraction Performance

Upon completion of processing the corpus of 1140 unique sentences on physical activity, 268 unique terms and concepts were structured into 33 semantic frames. Table 2 presents the performance of MUTT on identifying relevant physical activity terms and concepts obtained from the 2 rounds of evaluation. We manually reviewed the annotation results in MUTT. The result window of MUTT (see the Current Working Results area of Figure 1) allows users to review the terms and phrases captured by MUTT within the sentence that they belong to. Therefore, users can easily determine whether the recognized terms are relevant or any relevant terms are missed. The first round of evaluation was done with a collection of 100 unique sentences. MUTT reached an *F* score of 0.895 with this evaluation. After an additional training round, MUTT's performance was improved when evaluated with a second set of 240 unique sentences. With the *F* score of 0.950, the terms and concepts extracted using MUTT were deemed sufficiently comprehensive. No incorrect (ie, false positive) annotation was observed in either test set.

Physical Activity Ontology

The 268 unique terms and concepts identified with MUTT and 1 additional concept identified from the existing ontologies were structured into the PACO. PACO currently contains a total of 225 concept classes including 1 root class, which we label as PACO, and 9 defined classes. PACO contains 20 object properties (including 10 inverse properties) and 1 data property called *hasTotalAmountMin*. The main concept hierarchy is formed with 5 branches of *Activity*, *Exercise_effect*, *Exercise_equipment*, *Exercise_program*, and *Modifier.* Most of the prepared concepts were placed under the *Activity* and *Modifiers* branches, which are structured into multiple layers of classes. *Activity* is the largest branch that includes 2 primitive classes and 7 defined classes, which span to 4 subclass levels. The 2 primitive classes are *Daily_living_activity* that contains various household chores and home maintenance activities and *Exercise_leisure_activity* that includes various exercise, sports, and other hobbies involving physical activity such as dancing, fishing, and camping. Overall, 4 of the 7 defined classes are formed by exercise effects and include *Balance_exercise*, *Endurance_exercise*, *Flexibility_exercise*, and *Strength_exercise*. The remaining 3 classes are *Outdoor_leisure_activity*, *Water_sport*, and *Winter_sport*. These 7 defined classes support additional views that are commonly used to classify these physical activities.

The largest concept class is *Exercise_leisure_activity* under *Activity*, which contains 38 subclasses. The *Exercise_equipment* and *Exercise_program* branches are quite small and incomplete, and each contains only 7 and 3 subclasses, respectively. A number of expressions that represent various intensity levels such as "until sweat a lot and breathe hard" and "makes heart rate increase a bit" were instantiated under the 3 intensity-level classes of *Low, Moderate*, and *Vigorous*. Figure 3 shows the high-level hierarchy (asserted) of PACO. The PACO structural summarization metrics are presented in Table 3. PACO has been included in BioPortal [42].

Table 2. Multipurpose text processing tool annotation performance, and the number of sentences, terms and concepts used in the evaluation.

Test number	Sentences, n	Target terms and concepts, n	Recall	Precision	*F* score	Example terms or expressions missed
1	100	292	0.857	1.000	0.895	hike uphill, walk uphill, ride a bike, walk fast, exercise (generic), how many total hours, etc
2	240	443	0.940	1.000	0.950	push-ups, weightlifting, brisk walking, washing clothes by hand, calisthenics, squash, etc

Figure 3. Physical activity ontology high-level hierarchy.

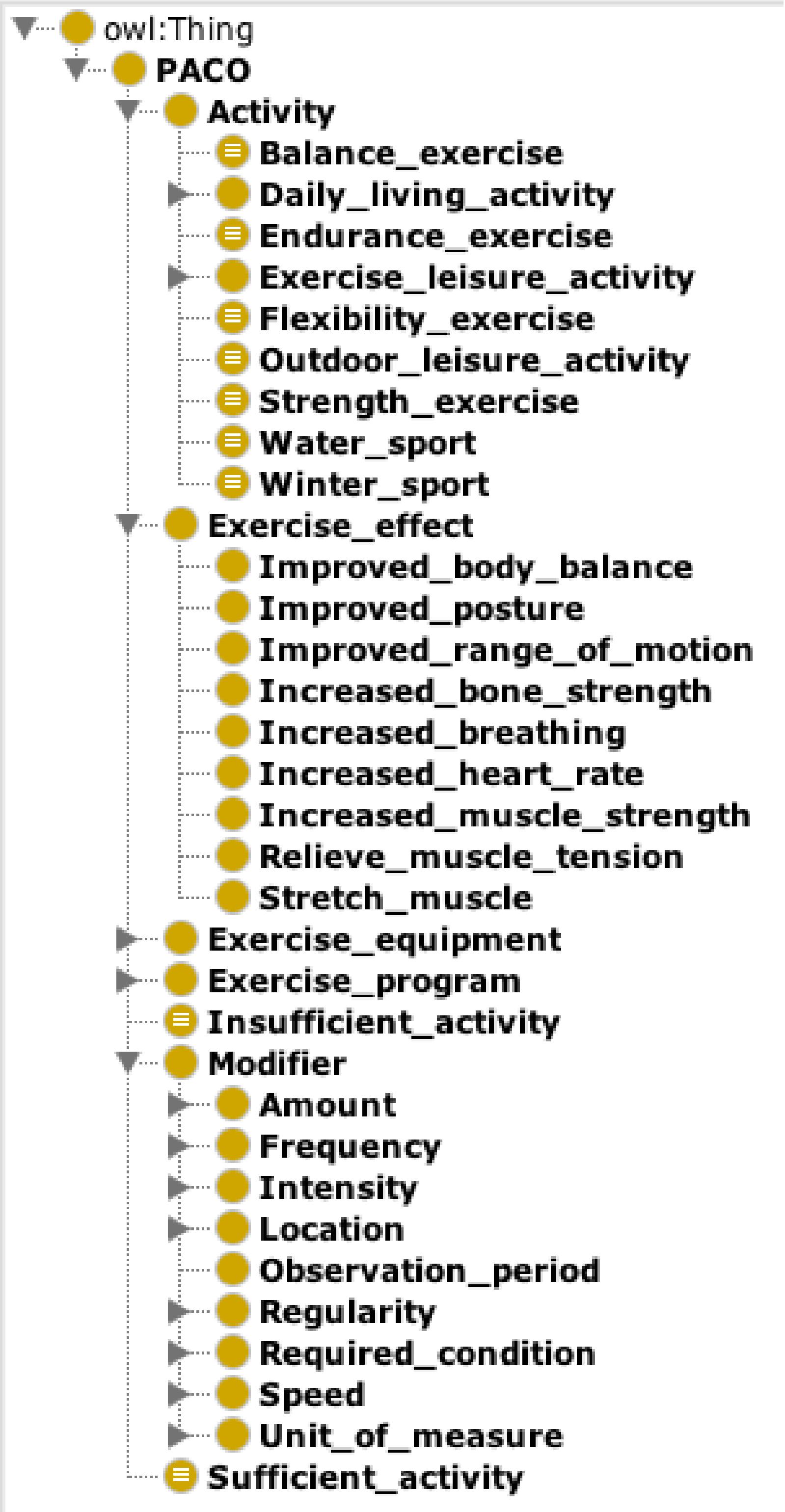


Table 3. The number of axioms and entities included in the Physical Activity Ontology.

Axioms and Entities in the Ontology	n
Class	225
Defined class	9
Maximum depth of classes	5
Minimum number of siblings	2
Maximum number of siblings	38
Logical axiom	587
Declaration axiom	297
Axiom for subsumption relations (subclass of)	397
Object property (excluding inverse property)	10
Data property	1
Instance (excluding 36 exercise statements)	23

Physical Activity Ontology Evaluation

PACO provided all activity types and intensity concepts that were required to represent the 36 exercise statements. The 36 exercise statements were all correctly classified into either the sufficient or insufficient level of exercise classification using FaCT++.

The results of the Protégé Ontology Debugger program indicated that PACO has well-defined concept classes, and all individuals are consistently instantiated under the relevant classes. However, OOPS! [39,40] caught a few issues in PACO. Absence of inverse object properties and absence of annotation of each class were identified as a minor problem. Interestingly, OOPS! recognized 3 pairs of potentially equivalent concepts—2 temporal concepts *Minute* and *Hour*, *Hockey* and *Field_hockey*, and *Hockey* and *Ice_hockey*. Not explicitly declaring their equivalence was identified as an important problem. OOPS! does not recognize the license information included as a PACO metadata, and absence of ontology license information was considered another important problem. We modified PACO to address these problems when possible. We added 10 inverse object properties that correspond to the 10 object properties. We annotated every class with internal concept identifier. We renamed Hockey to *Hockey_game* because ice hockey and field hockey are often called simply hockey in real-world communications.

Discussion

Principal Findings

We developed PACO as a conceptual foundation for systematically structuring and standardizing physical activity descriptions. PACO includes specific activity types and modifiers that are frequently used to further specify different properties of an activity. PACO was evaluated using the Ontology Debugger program of Protégé and the OOPS! program to ensure structural consistency and compliance to well-accepted ontology building principles.

Physical activities are often described with nontechnical terms and can be expressed in various forms that may include precise numeric measures (eg, walks 3 miles after dinner 1-2 times per week) to general colloquial descriptors (eg, occasionally go for a long walk after dinner). Existing biomedical terminology systems offer limited coverage for physical activity names and general descriptors. Instead of proposing to include more concepts and terms to the existing terminologies, we developed PACO, an ontology for a specialized scope of physical activities to reap the full benefit of an ontological approach to concept representation. For example, a collection of individual terms mapped to a standardized terminology does not capture the complete meaning as the semantic relations between the terms are not explicitly represented. PACO supports expressing complex concepts by linking physical activity concepts and modifier concepts using object properties based on the predefined semantics described in the ontology. In addition, activity names and types are classified from multiple perspectives in PACO using various classification criteria. These can be supported by concept post coordination with a compositional terminology that supports multiple inheritance. However, it is a nontrivial task to fully address diverse and complicated representational needs of physical activity descriptions for all potential queries that may originate from a broad scope of biomedical domains and applications.

PACO is one of the first ontologies that is dedicated to representing the concepts related to physical activity. PACO incorporates not only conceptual models but also individual concepts important to describe one's physical activity level such as activity types, intensity, and amount. PACO has a relatively simple asserted hierarchy where new concepts are easily added. To accommodate multiple ways of classifying physical activities, several defined classes were added and a multiple inheritance structure (ie, inferred hierarchy) was generated using a publicly available OWL-DL reasoner.

Limitations and Future Enhancement of Physical Activity Ontology

PACO contains concepts mainly derived from various assessment scales and questionnaires on physical activity. Therefore, it may have missed the concepts used in other types of physical activity descriptions found in various text sources such as patient exercise diary, clinical notes, and research articles. The *Activity* branch contains a number of specific types of physical activities including daily living activities and exercise/leisure activities that are frequently used to describe physical activity types in assessment scales and questionnaires. Although the *Activity* branch has the largest number of nodes and the deepest structure in PACO, it by no means contains the exhaustive set of activity type concepts. The *Activity* branch will continuously expand as more related sources of text are analyzed and incorporated. The multiple inheritance structure will also adaptively evolve as additional ways of classifying activities are identified.

Many physical activity questions employ phrases that describe physical responses to exercise such as "...[exercise] until you breathe harder than normal" or "...[exercise] makes you puff or pant," in addition to the general adjectives such as mild, moderate, strenuous, vigorous, etc. For example, 1 question of the Exercise Vital Sign uses an intensity descriptor of "... causes a light or heavy sweat," which indicates moderate or vigorous exercise. Although these "raw" expressions sound somewhat subjective, incorporating them into PACO was deemed important as exercise intensity can be a subjective experience influenced by people's age, overall health status, and fitness. These "raw" expressions were included in PACO as an instance of the intensity concept class Mild, Moderate, and Vigorous.

Intensity of exercise can also be captured by the type of activity itself. Many exercise guidelines and activity questionnaires provide specific types of activities as an example of indicating different exercise intensity levels. For example, brisk walking, water aerobics, and yoga are considered moderate activities, whereas jogging, aerobic dancing, and various competitive martial arts are considered vigorous activities [43]. In this version of PACO, named activities are not defined with an intensity level. However, we plan to attach the intensity property to the named activities in a revised PACO if it is deemed useful.

As an example-of-use demonstration, we classified 36 exercise habit statements into 2 clinically relevant exercise levels, that is, sufficient and insufficient, which are defined crudely based on the exercise amount and intensity. This is to show that once various exercise habit statements are formally represented using the concepts and properties in PACO, the formal representation can further be used to logically infer essential information (ie, whether the person is getting an adequate level of exercise or not). In reality, however, determining the adequacy of exercise level requires considering a person's individual characteristics such as demographics, body measures, health status, and overall physical fitness, in addition to the exercise description itself. A real-world application of algorithmically determining adequacy of one's exercise level will require incorporating objective measures such as metabolic equivalent of task-minutes (MET-minutes) [44] and heart-rate–based intensity measures (eg, 50%-70% increase of your maximum heart rate for moderate intensity) [45,46]. A future enhancement of PACO can consider including MET values as a property of specific named activities. In addition, exercise-induced heart-rate changes can be incorporated into the definitions of the exercise-level classes.

These limitations and the areas for future enhancements, however, do not detract from the motivating goal of PACO to contribute toward a precision medicine practice by facilitating the integration of heterogeneous data on physical activity generated by various sources. For example, PACO can provide standardized representations for a person's self-reported subjective descriptions on intensity and adequacy of physical activities. Moderate-intensity activities are considered to have the energy expenditure of 3 to 6 METs [44,47,48]. Walking 4.5 miles per hour falls in the moderate-intensity activity category, but to some people, this level of activity may be perceived as quite a high-intensity activity. Comparing these descriptions with the objective data collected from a person's mobile sensor device (ie, activity tracker) may lead to refined assessments and recommendations for improving a patient's physical activity level toward sufficiency [49].

Conclusions

Physical activity data are an important aspect to understanding general health, disease progress, and treatment outcome. The wide variety of ways of representing one's physical activity data has become a challenge with regard to analyzing them in conjunction with other clinical and biological data. As a first step toward standardizing and structuring heterogeneous descriptions on physical activity for integrative data analyses, we developed PACO with the concepts collected from physical activity scales and questionnaires. PACO was proven to be structurally consistent and cohesive and also demonstrated to be potentially useful in standardizing heterogeneous physical activity descriptions and classifying them into clinically meaningful categories that reflect adequacy of exercise. PACO will be continuously augmented to expand its concept coverage and semantic properties to support consistent documentation, standardization, and harmonization of physical activity data as described in various textual forms.

Acknowledgments

The authors thank Hedy Woo for helping with data preparation. This study was supported in part by the pilot grant from the Center for Nursing Research at Duke University School of Nursing.

Authors' Contributions

HK designed the study, prepared the data, built and evaluated PACO, and wrote the manuscript. JM prepared the data, built and evaluated PACO, and wrote the manuscript. RT prepared the data and wrote the manuscript. MUTT is the tool developed by RT for other studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions and sentences analyzed in this study.

[TXT File, 143KB - jmir_v21i4e12776_app1.txt]

References

1. Abel T. Measuring health lifestyles in a comparative analysis: theoretical issues and empirical findings. Soc Sci Med 1991;32(8):899-908. [Medline: 2031206]
2. Brugnara L, Murillo S, Novials A, Rojo-Martínez G, Soriguer F, Goday A, et al. Low physical activity and its association with diabetes and other cardiovascular risk factors: a nationwide, population-based study. PLoS One 2016 Aug 7;11(8):e0160959 [FREE Full text] [doi: 10.1371/journal.pone.0160959] [Medline: 27532610]
3. Husu P, Suni J, Vähä-Ypyä H, Sievänen H, Tokola K, Valkeinen H, et al. Objectively measured sedentary behavior and physical activity in a sample of Finnish adults: a cross-sectional study. BMC Public Health 2016 Dec 01;16:920 [FREE Full text] [doi: 10.1186/s12889-016-3591-y] [Medline: 27586887]
4. King AC, Oman RF, Brassington GS, Bliwise DL, Haskell WL. Moderate-intensity exercise and self-rated quality of sleep in older adults. A randomized controlled trial. J Am Med Assoc 1997 Jan 1;277(1):32-37. [Medline: 8980207]
5. King AC, Friedman R, Marcus B, Castro C, Forsyth L, Napolitano M, et al. Harnessing motivational forces in the promotion of physical activity: the Community Health Advice by Telephone (CHAT) project. Health Educ Res 2002 Oct;17(5):627-636. [Medline: 12408207]
6. Kokkinos PF, Narayan P, Papademetriou V. Exercise as hypertension therapy. Cardiol Clin 2001 Aug;19(3):507-516. [Medline: 11570120]
7. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795 [FREE Full text] [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
8. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
9. Asche CV, Seal B, Kahler KH, Oehrlein EM, Baumgartner MG. Evaluation of healthcare interventions and big data: review of associated data issues. Pharmacoeconomics 2017 Aug;35(8):759-765. [doi: 10.1007/s40273-017-0513-5] [Medline: 28474299]
10. Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): development and evaluation. J Clin Epidemiol 1993 Feb;46(2):153-162. [Medline: 8437031]
11. Rubinstein YR, McInnes P. NIH/NCATS/GRDR® Common Data Elements: a leading force for standardized data collection. Contemp Clin Trials 2015 May;42:78-80 [FREE Full text] [doi: 10.1016/j.cct.2015.03.003] [Medline: 25797358]
12. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, et al. The PhenX Toolkit: get the most from your measures. Am J Epidemiol 2011 Aug 01;174(3):253-260 [FREE Full text] [doi: 10.1093/aje/kwr193] [Medline: 21749974]
13. Logical Observation Identifiers Names and Codes. What LOINC is URL: https://loinc.org/get-started/what-loinc-is/ [accessed 2018-11-14] [WebCite Cache ID 73vSjNS63]
14. National Institute of Health Common Data Element Repository. NIH common data elements repository Internet URL: https://cde.nlm.nih.gov/ [accessed 2018-11-09] [WebCite Cache ID 73n1RqFQd]
15. Health Measures. Neuro-QoL URL: http://www.healthmeasures.net/explore-measurement-systems/neuro-qol [accessed 2018-11-12] [WebCite Cache ID 73sdpaX1B]
16. Saver JL, Warach S, Janis S, Odenkirchen J, Becker K, Benavente O, National Institute of Neurological DisordersStroke (NINDS) Stroke Common Data Element Working Group. Standardizing the structure of stroke clinical and epidemiologic research data: the National Institute of Neurological Disorders and Stroke (NINDS) Stroke Common Data Element (CDE) project. Stroke 2012 Apr;43(4):967-973 [FREE Full text] [doi: 10.1161/STROKEAHA.111.634352] [Medline: 22308239]
17. Institute of Medicine, Board on Population Health and Public Health Practice, Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records. Capturing Social And Behavioral Domains In Electronic Health Records: Phase 1. Washington, DC: National Academies Press; 2014.
18. Coleman KJ, Ngor E, Reynolds K, Quinn VP, Koebnick C, Young DR, et al. Initial validation of an exercise "vital sign" in electronic medical records. Med Sci Sports Exerc 2012 Nov;44(11):2071-2076. [doi: 10.1249/MSS.0b013e3182630ec1] [Medline: 22688832]
19. Dou D, Wang H. Bioportal. 2015. SMASH Ontology URL: https://bioportal.bioontology.org/ontologies/SMASH [accessed 2018-11-09] [WebCite Cache ID 73n0z6Tc5]

20. Silva P, Andrade MT, Carvalho P, Mota J. A structured and flexible language for physical activity assessment and characterization. J Sports Med (Hindawi Publ Corp) 2013;2013:420916 [FREE Full text] [doi: 10.1155/2013/420916] [Medline: 26464878]
21. Bamparopoulos G, Konstantinidis E, Bratsas C, Bamidis PD. Towards exergaming commons: composing the exergame ontology for publishing open game data. J Biomed Semantics 2016 Feb 9;7:4 [FREE Full text] [doi: 10.1186/s13326-016-0046-4] [Medline: 26865947]
22. Hi5Living. 2015. Healthy living questionnaire: Self assessment URL: http://www.hi5living.org/self-assessments/ [accessed 2018-11-09] [WebCite Cache ID 73n19cfUg]
23. Topolski TD, LoGerfo J, Patrick DL, Williams B, Walwick J, Patrick MB. The Rapid Assessment of Physical Activity (RAPA) among older adults. Prev Chronic Dis 2006 Oct;3(4):A118. [Medline: 16978493]
24. Smith BJ, Marshall AL, Huang N. Screening for physical activity in family practice: evaluation of two brief assessment tools. Am J Prev Med 2005 Nov;29(4):256-264. [doi: 10.1016/j.amepre.2005.07.005] [Medline: 16242587]
25. Cohen KB, Verspoor K, Johnson HL, Roeder C, Ogren PV, Baumgartner WA, et al. High-precision biological event extraction: effects of system and of data. Comput Intell 2011 Nov;27(4):681-701 [FREE Full text] [doi: 10.1111/j.1467-8640.2011.00405.x] [Medline: 25937701]
26. Zhao Y, Fesharaki NJ, Liu H, Luo J. Using data-driven sublanguage pattern mining to induce knowledge models: application in medical image reports knowledge representation. BMC Med Inform Decis Mak 2018 Dec 06;18(1):61 [FREE Full text] [doi: 10.1186/s12911-018-0645-3] [Medline: 29980203]
27. Taira RK, Taira R, Bui A, Bui AA, Hsu W, Bashyam V, et al. A tool for improving the longitudinal imaging characterization for neuro-oncology cases. AMIA Annu Symp Proc 2008 Nov 06:712-716 [FREE Full text] [Medline: 18999139]
28. Taira RK, Soderland SG, Jakobovits RM. Automatic structuring of radiology free-text reports. Radiographics 2001;21(1):237-245. [doi: 10.1148/radiographics.21.1.g01ja18237] [Medline: 11158658]
29. Hsu W, Taira RK. Tools for improving the characterization and visualization of changes in neuro-oncology patients. AMIA Annu Symp Proc 2010 Nov 13;2010:316-320 [FREE Full text] [Medline: 21346992]
30. Kim H, Kim J, Shenvi E, Quach J, Sutjiadi B, Richardson A, et al. Developing a semantic model to describe physical activity data. Stud Health Technol Inform 2016;225:447-451. [Medline: 27332240]
31. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 2009 Jul;37(Web Server issue):W170-W173 [FREE Full text] [doi: 10.1093/nar/gkp440] [Medline: 19483092]
32. BioPortal. 2009 Jul. SNOMED CT URL: https://bioportal.bioontology.org/ontologies/SNOMEDCT [accessed 2019-03-14] [WebCite Cache ID 76rfzCMH1]
33. SNOMED International. SNOMED-CT URL: http://www.snomed.org/ [accessed 2018-10-29] [WebCite Cache ID 76pBYlbub]
34. BioPortal. National Cancer Institute Thesaurus URL: https://bioportal.bioontology.org/ontologies/NCIT [accessed 2018-10-29] [WebCite Cache ID 76pBpiv5r]
35. NCI Term Browser. NCI theasurus URL: https://ncit.nci.nih.gov/ncitbrowser/ [accessed 2019-03-12] [WebCite Cache ID 76pC36I7e]
36. NHS Digital. 2014. Read Codes URL: https://digital.nhs.uk/services/terminology-and-classifications/read-codes [accessed 2019-03-12] [WebCite Cache ID 76pCGEW42]
37. OWL@Manchester. FaCT++ reasoner URL: http://owl.cs.manchester.ac.uk/tools/fact/ [accessed 2018-11-09] [WebCite Cache ID 73n25IBHr]
38. Schekotihin K, Rodler P, Schmid W, Horridge M, Tudorache T. Test-driven Ontology Development in Protégé. In: International Conference on Biological Ontology 2018. 2018 Jun 8 Presented at: Proceedings of the 9th International Conference on Biological Ontology (ICBO 2018); 2018; Corvallis, OR URL: http://icbo2018.cgrb.oregonstate.edu/node/134
39. Poveda-Villalón M, Gómez-Pérez A, Suárez-Figueroa M. OOPS! (OntOlogy Pitfall Scanner!). Int J Semant Web Inf Syst 2014;10(2):7-34. [doi: 10.4018/ijswis.2014040102]
40. OOPS! OntOlogy Pitfall Scanner!. URL: http://oops.linkeddata.es/ [accessed 2018-10-21] [WebCite Cache ID 76pE02FDD]
41. Health.gov. 2008. Physical activity guidelines for Americans URL: https://health.gov/paguidelines/ [accessed 2018-11-09] [WebCite Cache ID 73n1w5aZP]
42. BioPortal. 2019. Physical Activity Ontology URL: https://bioportal.bioontology.org/ontologies/PACO [accessed 2018-11-12] [WebCite Cache ID 73sfO5XIa]
43. Centers for Disease Control and Prevention. General Physical Activities Defined by Level of Intensity URL: https://www.cdc.gov/nccdphp/dnpa/physical/pdf/PA_Intensity_table_2_1.pdf [accessed 2018-10-21] [WebCite Cache ID 73n2cnGXA]
44. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
45. Karvonen J, Vuorimaa T. Heart rate and exercise intensity during sports activities. Practical application. Sports Med 1988 May;5(5):303-311. [doi: 10.2165/00007256-198805050-00002] [Medline: 3387734]

46. Mayo Clinic. 2018. Exercise intensity: How to measure it Internet URL: https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887 [accessed 2018-10-29] [WebCite Cache ID 76pF1T8YY]
47. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
48. Piercy K, Troiano R, Ballard R, Carlson S, Fulton J, Galuska D, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Dec 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
49. Franklin BA, Brinks J, Berra K, Lavie CJ, Gordon NF, Sperling LS. Using metabolic equivalents in clinical practice. Am J Cardiol 2018 Feb 01;121(3):382-387. [doi: 10.1016/j.amjcard.2017.10.033] [Medline: 29229271]

Abbreviations

CDE: common data elements
DL: Description Logic
EHR: electronic health record
IOM: Institute of Medicine
MET: metabolic equivalent of task
MUTT: Multipurpose Text Processing Tool
NLP: natural language processing
ONC: Office of National Coordinators
OOPS: Ontology Pitfall Scanner
OPA: Ontology for Assessing Physical Activity and Sedentary Behavior
OPE: Ontology of Physical Exercise
OWL: Web Ontology Language
PACO: Physical Activity Ontology
PhenX: Phenotypes and eXposures
SMASH: Semantic Mining of Activity, Social, and Health Data

Edited by M Focsa; submitted 14.11.18; peer-reviewed by C Bratsas, P Santaguida, F Li; comments to author 04.02.19; revised version received 15.02.19; accepted 04.03.19; published 23.04.19.

Please cite as:
Kim H, Mentzer J, Taira R
Developing a Physical Activity Ontology to Support the Interoperability of Physical Activity Data
J Med Internet Res 2019;21(4):e12776
URL: http://www.jmir.org/2019/4/e12776/
doi:10.2196/12776
PMID:31012864

Original Paper

“It’s Not Just Technology, It’s People”: Constructing a Conceptual Model of Shared Health Informatics for Tracking in Chronic Illness Management

Lisa M Vizer[1], PhD; Jordan Eschler[2], PhD; Bon Mi Koo[1], PhD; James Ralston[3], MD; Wanda Pratt[4], PhD; Sean Munson[4], PhD

[1]Division of General Medicine and Clinical Epidemiology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[2]Northwestern University, Chicago, IL, United States
[3]Kaiser Permanente Washington, Seattle, WA, United States
[4]University of Washington, Seattle, WA, United States

Corresponding Author:
Lisa M Vizer, PhD
Division of General Medicine and Clinical Epidemiology
School of Medicine
University of North Carolina at Chapel Hill
Health Sciences Library, Rm 333A, CB 7110
335 S Columbia St
Chapel Hill, NC, 27599-7110
United States
Phone: 1 4109082774
Email: LMVizer@unc.edu

Abstract

Background: For many people, tracking health indicators is central to managing a chronic illness. However, previous informatics research has largely viewed tracking as a solitary process that lacks the characteristics essential to tracking in support of chronic illness management.

Objective: To inform development of effective technologies that aid tracking of health indicators to support chronic illness management, this study aimed to construct a health informatics model that accurately describes the work and social context of that tracking work.

Methods: As part of a larger project, we conducted semistructured interviews with 40 adults concerning their chronic illness management practices, including tracking and communication. We also assembled transcripts of 30 publicly available videos of 24 adults discussing tracking processes for managing their own chronic illness. We used qualitative methods to analyze interviews and video transcripts through the lens of ongoing personal and health informatics research.

Results: We have described the people and work involved in tracking in support of chronic illness management and contributed a Conceptual Model of Shared Health Informatics (CoMSHI). Specifically, we identified the need for a health informatics model that (1) incorporates the ongoing nature of tracking work and (2) represents the social dimension of tracking for illness management. Our model depicts communication, information, collection, integration, reflection, and action work in the social context of the person with chronic illness, informal carers, health care providers, and community members.

Conclusions: The resulting CoMSHI yields a more detailed and nuanced viewpoint of tracking in support of chronic illness management and can inform technology design to improve tracking tools to support people in more confident and capable chronic illness management.

(J Med Internet Res 2019;21(4):e10830) doi:10.2196/10830

KEYWORDS

consumer health informatics; chronic illness; patient generated health data; patient reported outcomes; workflow; information seeking behavior; shared decision making

Introduction

Background

Chronic illness, defined as “tend[ing] to be of long duration and are the result of a combination of genetic, physiological, environmental and behavioural factors” [1], is the leading cause of poor health, disability, and death, accounting for up to 86% of health care spending [2]. Over half of Americans have a chronic illness, with 25% having more than one [3]. To live well with a chronic illness, people must engage effective management strategies, including tracking behaviors, biometrics, and symptoms, within the constraints of their everyday lives [4,5], and they are turning to technology solutions in increasing numbers [6] (we refer to *people* rather than *patients* throughout this paper to acknowledge the whole person who has a chronic illness and not just their patient role, which may not resonate outside of the clinical setting; management activities take place, overwhelmingly, in the course of day-to-day living [7,8]).

Many health informatics tools—such as continuous glucose monitors, activity trackers (eg, Fitbit), heart rate monitors, and smartphone apps (eg, OnTrack)—promise to ease the work of tracking in support of chronic illness management. However, many people do not find these tools useful [9], and those who do use these tools often experience substantial barriers to effective use [10,11]. Despite these barriers, many people managing chronic illness must regularly track health indicators to maintain or improve health [6,12], whether using a digital or analog tool, and need better support for successful tracking practices. Understanding the processes related to the tracking work of people managing chronic illness is critical for developing health informatics tools that support confident, capable, and effective illness management.

As noted by health informatics researchers [13,14], however, current informatics models do not adequately represent the social context or the tracking tasks undertaken to manage chronic illness. A model accounting for the people and work involved in tracking can guide the design of technologies and services that better integrate into people’s lives and support their goals. To bridge this gap in representation, we developed an improved health informatics model through a qualitative analysis of 69 transcripts from 63 people describing their chronic illness management routines. The contribution of this study is a Conceptual Model of Shared Health Informatics (CoMSHI) describing tracking in support of chronic illness management. We delineate the components of the CoMSHI and relationships among those components with the goal of supporting the design of informatics tools that align with chronic illness tracking work [8,13,14].

Related Literature

Our study builds on and extends the literature on tracking behavior in general, tracking for chronic illness management, and current personal and health informatics models.

Tracking

People engage in tracking activities for a wide range of purposes, from understanding finances to improving productivity to supporting artistic expression [10]. At the forefront of tracking technology and practice is the Quantified Self (QS) community [15], an enthusiastic group of trackers who describe themselves as an “international collaboration of users and makers of self-tracking tools.” QS members are active around the world with regional meetings and an annual conference. At meetings, trackers give *Show and Tell* presentations to describe their experiences.

Researchers study the practices of these and other trackers to understand their collection and sense-making strategies [10] and develop models of the tracking process [11,16] (discussed later in this section). People track using a variety of methods including automatic methods, such as bank logs, and manual methods, such as calendars [11]. Tools may be analog, such as paper journals, or digital, such as smartphone calendars [11]. After collecting data, people reflect on relationships among and patterns in data. Reflection can lead to insights about behavior or decisions to change behavior. However, barriers to successful tracking include difficulties with (1) deciding what data to collect and what tools to use, (2) using tools, (3) collating and formatting data, (4) understanding and reflecting on the data, and (5) formulating action plans based on the data [10,11].

Tracking for Management of Chronic Illness

Some QS members track to identify and solve health problems, although most health-related tracking is for general wellness. Similarly, much research focuses on tracking for general health and wellness (eg, [17-21]). Tracking has become an important and prevalent activity among the general public; the Pew Internet & American Life Project [6] recently reported that 69% of Americans track health data for themselves or a loved one.

However, many people managing a chronic illness must monitor their symptoms and their health through tracking. As of 2013, 19% of adults with no chronic illness tracked a health factor, whereas 40% of those with 1 chronic condition engaged in tracking, and 62% of adults with 2 or more chronic illnesses track one or more health factors [6]. This tracking can be an effective part of managing chronic illness, improving health outcomes [22], and communicating with health care providers [23]. Studies examining the lived experiences of tracking find that it is often coordinated or influenced by communication with health experts (eg, [24-26]), peers (eg, [14,20,21]), family members (eg, [27-29]), and colleagues and workplace programs (eg, [30]). Although much of the literature on tracking acknowledges that these practices happen in and are influenced by various social contexts, the models that guide design and study of tracking tools focus on individual tracking [11,16].

As more people managing chronic illness participate in health tracking, health care providers and researchers need to understand their work and social ecosystem. Moreover, although technology facilitates tracking, limited evidence supports the efficacy of specific tools to accomplish successful illness management [31]. Barriers to successful tracking for health include (1) insufficient support for collaboration with a provider [32], (2) difficulty making sense of data leading to lapses in tracking [16], and (3) difficulty remembering to track or deciding what to track [33]. Furthermore, many studies find that the apps available to support tracking for chronic illnesses are of poor quality [34-40], pointing to a lack of understanding of the needs of people managing chronic illness.

MacLeod et al [24] interviewed 12 people with a range of chronic illnesses who tracked some aspect of their health. They found that people wished to understand how their illnesses affected their lives within the context of information from health care providers. Mamykina et al [41] developed and tested MAHI—a mobile tracking and communication tool for people with diabetes. Their work revealed that, even with coaching from a diabetes educator, people newly diagnosed with diabetes struggled to develop self-efficacy and reflective thinking skills with regard to the data they captured. Our study sought to complement these studies by contributing an understanding of the processes by which trackers engage with a social dimension of tracking behavior.

Personal and Health Informatics Models

A total of 4 informatics models have articulated the tracking process: Li et al [11]; Epstein et al [16]; Swan [42]; and Murnane et al [43]. Li et al [11] and Epstein et al [16] focus on general personal informatics, whereas Swan [42] and Murnane et al [43] discuss informatics in the context of health.

Li et al [11] conducted an interview study of 11 self-trackers to derive a stage-based personal informatics model. Their model describes the process of general self-tracking for any purpose and is composed of *preparation*, *collection*, *integration*, *reflection*, and *action* stages. This model focuses solely on the stages through which a single person progresses in self-tracking. The authors highlight the dependencies between each stage: a mistake in preparation can cause someone to collect the wrong data, and these problems then cascade to the integration, reflection, and action stages. Whooley et al [44] expand on the integration stage of Li et al’s personal informatics model with a discussion of *why* people track and *how* they integrate their data. More recently, Mamykina et al [26] studied the process of self-discovery in a structured diabetes education program, showing how tracking can scaffold learning and reflection for diabetes management.

The stage-based personal informatics model describes an ideal process for tracking, but that process can break down when it encounters the realities of everyday life [45]. To describe general tracking in everyday life, Epstein et al [16] propose the Lived Informatics model. This model is based on 184 surveys and 22 interviews on self-tracking behaviors for physical activity, location, and finances. The authors refined Li et al’s model by dividing its original *preparation* stage into 2 stages: *deciding* and *selecting*. Introducing a cycle named *tracking and acting*, they describe an iterative progression through *collection*, *integration*, and *reflection*. Finally, this model anticipates that people will lapse in their tracking practice either temporarily or permanently.

Swan [42] proposes a model for Patient-driven Health Care that includes self-tracking among the actions in which a patient might engage. The patient is the only one engaging in work, and people other than the patient are included only peripherally. Furthermore, the evidence base for developing the model is unclear.

Murnane et al [43] examined applying Ecological Systems Theory [46] to the work of tracking in long-term management of severe mental illness [43]. The resulting model describes the influences and resources available to people, including close personal ties, indirect institutional influences, and an individual’s sociocultural context. They discuss how personal informatics tools and data form an *informatics layer* that can mediate interactions with these other services, though the model does not articulate activities in the tracking process.

Although researchers have used these models as a lens through which to study tracking related to health [14,47], they found that they lack elements important to the health context. Costa Figueirido et al [14] studied women making sense of infertility and found that the stage-based model did not adequately represent the fluidity of work or the collaboration in which women engaged. Mishra et al [47] studied people tracking while hospitalized and also observed that the stage-based model did not characterize the collaboration occurring around data. Costa Figueirido et al [14], Mishra et al [47], and Valdez and Brennan [13] all explicitly indicate a need for a model that more closely aligns to the unique needs of the health context.

This study addresses key gaps by constructing a model representing the work and people involved in tracking to support chronic illness management.

Methods

Datasets

We collected data and conducted interviews with people managing chronic illness in the following 3 groups: QS speakers presenting about managing chronic illness, adults managing type 2 diabetes, and mothers managing a child’s asthma. We chose publicly available QS videos to access a sizeable sample of expert trackers using innovative technology, and we chose to interview people in the community to gain the perspective of more typical trackers and technology users.

We obtained institutional review board approval from Group Health Research Institute, University of Washington, and University of North Carolina at Chapel Hill for collection, analysis, and reporting of data used for this study.

Quantified Self Cohort

For the QS dataset, we selected videos posted publicly on the QS blog [15] between January 2012 (blog inception) and December 2018 (end of data collection) that focused on managing a chronic illness, as defined in the Introduction. QS speakers are enthusiastic technology and tracking hobbyists who often characterize their tracking practices as innovative. They share their experiences at local and worldwide meetings to disseminate information about their routines and insights. In these videos, people described their work processes: information they tracked, how they analyzed and learned from that information, and when and how they shared information with others.

These presentations are meant to instruct other QSs, therefore giving us access to their expertise similar to an interview elicitation. Given their proficiency, we expected QS presentations to give us an overview of the cutting edge of how people build knowledge about chronic illness through tracking. As the videos are publicly accessible, we were not required to obtain consent for their use. Although presenter names are included on the QS blog, we chose to use anonymous identifiers for analysis and reporting.

Interview Cohorts

The interview participants were patients of primary care providers in Group Health owned and operated clinics (now Kaiser Permanente Washington), a large integrated health care delivery system in Washington State providing care to over 300,000 people. We recruited adults managing type 2 diabetes and mothers managing asthma for at least one child aged 12 years or younger. These 2 diagnoses were chosen because each requires daily health-related tasks and frequent contact with health care providers and health care systems (eg, scheduling appointments, filling prescriptions, scheduling lab tests, and asking questions outside of formal appointments). We used purposeful sampling to identify participants representative of the general population in the Northwest United States based on gender, ethnicity, technology use (with recorded use of a patient portal as a proxy), and education. All interview participants completed an informed consent process. We conducted semistructured interviews in each participant's home inquiring about health goals, priorities for completing health tasks, and workflow in attaining those goals and tasks. The workflows articulated by participants included information on tracking and communication in support of health management. Group Health Research Institute contracted a *Health Insurance Portability and Accountability Act* (HIPAA)–approved outside agency to transcribe and redact audio recordings of interviews.

Analysis

We coded transcripts with the ATLAS.ti (Scientific Software Development GmbH) software package using open coding. Our analysis was informed by tracking and personal informatics literature (to ensure construct validity) as well as themes we identified on initial read-throughs regarding tracking behavior.

Specifically, 3 authors (LMV, JCE, and BK) iterated through a subset of the data corpus, revising the list of open codes to refine the scope of the analysis and clearly define individual codes. When all coders agreed that the list of open codes sufficiently represented the themes related to the scope of the inquiry (ie, tracking behaviors and social dimensions of such behaviors), the coders converted the list of open codes to an axial coding scheme using affinity diagramming [48]. We then applied the axial codes to the complete data corpus. The resulting coded dataset, and the process of tracking that it described, was used to define a new conceptual model that describes the work and people involved in tracking to manage chronic illness. The following section first describes the analysis results then describes the new conceptual model.

Results

Datasets

Our dataset included transcripts of videos and interviews with people managing chronic illness, and the analysis guided the definition of the components of the tracking process supporting chronic illness management and construction of a model of this process. We collected data from 64 people managing a chronic illness. Data came from 24 QS speakers, 20 adults with type 2 diabetes, and 20 parents of a child with asthma. The QS cohort consists of highly proficient trackers and technology users, in contrast to the interview cohorts made up of mainstream trackers and technology users.

Quantified Self Cohort

From among the videos on the QS blog, we identified 30 publicly available videos meeting inclusion criteria, with a total running time of over 6 hours and 57 min. Videos were from 24 people; 4 speakers made 2 presentations each, and 1 speaker made 3 presentations. A total of 16 speakers were male (67%). One speaker acted as an informal carer (4%) for her child (ie, an unpaid provider of health-related care and support).

All speakers appeared to be of non-Hispanic white race and ethnicity and therefore do not represent the demographics of the general population. However, we judged that the videos still provide valuable insights, and the homogeneity of this sample is somewhat balanced by the diversity of the interview samples discussed in the next section. Table 1 describes the speakers' gender, race, tracking role, employment, and diagnosed illnesses. The mean tracking interval that the speakers referenced was 3 years. The average video length was 13 min 50 seconds. QS videos are denoted with Q# identifiers in the quotes highlighted in the results.

Table 1. Quantified Self (QS) speaker demographics.

Demographics	n (%)[a]
Gender	
Female	8 (33)
Male	16 (67)
Race and ethnicity	
White (non-Hispanic)	24 (100)
Role	
Person with chronic illness	23 (96)
Carer of person with chronic illness	1 (4)
Employment	
Technology industry	8 (33)
Other industry	4 (17)
Academia	4 (17)
Health care (eg, physicians and nurses)	3(13)
Not reported	5 (21)
Illness	
Diabetes	9 (38)
Allergies (food or environmental)	4 (17)
Parkinson disease	2 (8)
Crohn disease	2 (8)
Arrhythmia	1 (4)
Chronic fatigue	1 (4)
Chronic headaches	1 (4)
Chronic neurological Lyme disease	1 (4)
Heart valve disorder	1 (4)
Panic disorder	1 (4)
Restless leg syndrome	1 (4)

[a] Percentages are rounded to the nearest whole number.

Interview Cohorts

We enrolled 20 adults with type 2 diabetes and 20 mothers of children aged 12 years and older with asthma from among the patients of Group Health clinics. Table 2 describes participant demographics. Interviews ranged in length from 45 to 90 min with an average of about 60 min. For diabetes cohort participants with an informal carer, we invited them to participate in the interviews if possible. Diabetes cohort participants have D# identifiers in the quotes highlighted in the results. Asthma cohort participants are denoted by A# identifiers.

Analysis

One author (LMV) transcribed the QS videos, and a HIPAA-approved vendor transcribed the interviews. We analyzed all transcripts using the qualitative analysis method outlined in the Methods section. We have discussed the tracking components we observed and described the model we constructed from these components.

Components of Tracking in Support of Chronic Illness Management

The primary themes emerging from our analysis consist of 2 parts—*actors* and *work*. The types of actors and work are summarized in Textbox 1.

We have discussed each actor and type of work, supported with examples from our analysis.

Actors

Actors are the *person with chronic illness*, *informal carers*, *health care providers*, and *community members*. These actors interact with each other and can all perform aspects of work, as described below. This definition extends beyond people included in the models of personal and health informatics from Li et al [11], Epstein et al [16], and Swan [42].

Table 2. Interview cohort demographics.

Demographics	Asthma	Diabetes
Gender, n (%)		
Female	20 (100)	10 (50)
Male	0 (0)	10 (50)
Age (years), mean	37.5	64.5
Education, n (%)		
High school or less	4 (20)	8 (40)
At least some college	16 (80)	12 (60)
Race and ethnicity, n (%)		
Asian	0 (0)	2 (10)
Black	6 (30)	6 (30)
White (non-Hispanic)	10 (50)	10 (50)
Other or no ethnicity given	4 (20)	2 (10)
Hispanic[a]	1 (5)	1 (5)

[a] Hispanic ethnicity designation overlapped with other designations of race.

Textbox 1. Types of actors and work identified through transcript analysis.

Model Components and Types

- Actors
  - Person with chronic illness, optionally including informal carers
  - Health care providers
  - Community members
- Work
  - Communication
  - Information
  - Collection
  - Integration
  - Reflection
  - Action

Person With Chronic Illness and Informal Carers

The *person with chronic illness* and *informal carers* are the actors most affected by the success or failure of chronic illness management and are therefore central to the tracking process. Informal carers are usually an unpaid spouse, partner, adult child, or parent. Carers often actively include the person with chronic illness in tracking work and may act as advocates or facilitators in managing chronic conditions. Carer involvement—which is crucial (eg, legally or financially) in certain situations, such as a parent advocating and caring for a child with chronic illness—is one example of the fundamentally social nature of tracking in support of chronic illness management and is not adequately described by previous informatics models. To this point, 1 carer was the mother of a child with type 1 diabetes, who told her son, “you’re a scientist along with us, and you’re making these discoveries” (Q25). Other research describes the dynamics between the person with chronic illness and informal carers in more depth, especially with regard to patient portals [28,29].

Health Care Providers

Health care providers are skilled health professionals involved in a person’s care. Although most people mention physicians when talking about health care providers, our analysis also noted many types of nurses (eg, school nurses, nurse practitioners, and homecare nurses), physical therapists, pharmacists, nutritionists, and others. This is consistent with other literature on chronic illness care [49].

Community Members

Community members are nonhealth professionals with whom other actors interact. This definition is more inclusive than Swan’s, which included only peers. This actor includes the widest variety of people, such as intimate partners, friends, roommates, others with chronic illness, colleagues, or schoolteachers. Other literature describes the community of a person with chronic illness more in depth [13,50-53].

Work

The types of tracking *work* include *communication*, *information*, *collection*, *integration*, *reflection*, and *action*. These types of work are similar to the stage-based model [11] but add *communication* work (to incorporate interactions between actors) and redefine *preparation* to *information*. Perhaps most important for the mechanics of work processes, we observed that *unconstrained transitions* described the structure of tracking for chronic illness management better than discrete stages. This reflects both the continuous *and* social natures of work revealed in our datasets. As also described by Epstein et al [16] and Costa Figueirido et al [14], our analysis showed that different types of work can occur simultaneously. Furthermore, any actor can engage in any work, and actors often collaborate or hand off work. We have discussed each type of work, dependencies, and workflow.

Communication

Communication work encompasses interactions between actors. These interactions may involve illness-related information, tracked data, visualizations, or motivational support. This work is particularly important in management of chronic illness because of the number of actors and amount of work involved. Valdez et al [8] refer to this as *articulation work*.

People with chronic illness and carers regularly manage communication tasks with others, often leveraging others' expertise. One mother who we interviewed had a friend who helped her better understand her child's allergy triggers:

> *We went to a friend's house and they had a dog, and my friend's a doctor and she was like "you know, she's having some kind of reaction to something, what's going on?"* [A4]

Some people struggled with lack of technology support for communication. Although she faithfully uses tracking to help her manage type 2 diabetes, Quantified Self speaker Q30 wishes she could easily share her data with her physician and family:

> *I would like an option to share these data points with my primary care provider so that he can see that I'm doing well and feeling well and my numbers are reflecting that. Also I'd like to be able to share this with my family, especially as I get older.* [Q30]

Information

Information work describes an ongoing process of accumulating information to support tracking. This type of work is most analogous to Li et al's [11] *preparation* stage, but we found that rather than engaging in just preparation, actors worked to accumulate a body of knowledge regarding aspects of tracking and illness. They used a wide variety of sources including other actors and third-party information sources, such as Wikipedia or medical websites, to learn terminology, make decisions, and understand feedback and outcomes (Q1). People with chronic illness and carers perform much of the information work, as it is specific to the individual's experience of the illness.

In contrast to the preparation work described in Li et al's model [11], information work informs people throughout the tracking process. Information can come from communication with other actors, such as health care providers:

> *I went and talked to my doctor about restless leg. We had a nice discussion about the genomic, the genetic aspects of this. He had some website stuff to go to.* [Q9]

One QS speaker describes doing information work while investigating patterns in her nutrition and symptom data. She engaged in this work simultaneously with reflection:

> *I got suspicious of bell pepper, tomatoes, and eggplant...It turns out they're in the same family. It's called nightshade. It has neurotoxins in it. They inhibit cholinesterase. What does cholinesterase do? Oh, my word. This...looks like what's been happening.* [Q11]

As described by Valdez et al [8] in their patient work framework, *Information* and *Communication* work support the rest of the work of tracking.

Collection

Collection work involves data-gathering activities. Actors use tools (eg, glucometer, blood pressure cuff, journal, spreadsheet, and smartphone) to collect data (eg, numeric, text, or picture; objective or subjective) depending on the illness and health goals. Objective data include blood glucose levels, blood pressure, peak flow meter readings, geographic locations, and food intake. Subjective data include discomfort levels and degree of breathing difficulty.

Although most people recorded data in text or numeric form, some people used photos and video. These rich data types convey more information than a simple number and can be especially helpful in tracking food intake or changes in movement over time (Q20 and Q24). Speaker Q14 even used the quality of her handwriting in her headache journal to corroborate headache severity. Chung et al [25] similarly discuss types of data as well as boundary-negotiating artifacts generated through tracking.

Collection can also be collaborative, particularly in families [27]. For example, some parents of children with asthma share collection duties (eg, A6), especially if they share custody. Furthermore, spouses with similar conditions may track together (D6), and carers may track in collaboration with the people they support (Q25).

Integration

Integration work involves transforming data for analysis. People detailed the ways they collated and displayed data, with most people using a simple spreadsheet and graphs. Q10 describes how he visualizes the sneezes that are a symptom of his allergies:

> *This is a different way of looking at my sneezes. It's a cumulative graph...and the slope indicated how fast I produce sneezes. So if it's flat I don't produce as*

> *many sneezes and if it's very steep I produce a lot of sneezes in a short amount of time.* [Q10]

Integration work is usually performed to more deeply understand interactions between types of data, such as the effect of medication or treatment regimens on specific symptoms or the effect of stress on blood glucose levels. D4 showed his integration work for weight change and medication intake:

> *This is my chart that I made. I went into Excel...This is my weight. I weigh myself every day. See, I gained a couple pounds overnight...I'm going to have to...make sure I take three [medications] in the morning and three at night.* [D4]

Some people with chronic illness and carers use patient portals to make charts or tables with their data. A17 explained that she used her patient portal to integrate data:

> *I can chart my progress. I can see if my numbers are going up or going down, I can see my blood pressure. It's not a test, but it's on there and I can see what my blood pressure was when I went in for the visit.* [A17]

Many people drew inspiration to continue tracking from the visualizations they produced. Q17 described the information visualizations she used as "incredibly motivating."

Reflection

Reflection work represents time spent engaging with data, making meaning from data, and considering the tracking experience itself. People with chronic illness or carers are usually primarily involved in reflection, with health care providers and community members providing additional insight. On the basis of an outcome, actors may decide to make adjustments or do something new. D8 reflected on how food intake affected her blood glucose:

> *I was writing down everything I ate during the...day and looking at the difference in my blood sugar, what caused it to be higher, and I had everything right there so that was more helpful.* [D8]

We also observed collaborative reflection work, consistent with previous research [14]. D1 discussed working with her pharmacist:

> *The pharmacist got involved in my cholesterol medication. She wanted me to go up a dose so we did a lot of communicating that way and that worked out.* [D1]

For an in-depth discussion of coordinated reflection, see Schroeder et al [32] discussing the work of people with Irritable Bowel Syndrome.

Action

Action work describes steps people take based on reflection or information work, often in collaboration with others. Some people talked about making incremental adjustments to their daily routine (Q19), but other speakers were inspired to make more substantial lifestyle changes, such as avoiding a medication (Tylenol for Q19) or cutting out foods (eliminating caffeine for Q3 to reduce panic attacks). Some also described weighing evidence from information and reflection work to synthesize conflicting advice from health care providers and decide on a plan of action (Q23).

A14 explained her new strategy for organizing medication after reflecting on gaps in medication logs:

> *I split them all up, and he was there [at his dad's house] for two weeks so I bought several of these [pill organizers] because he takes one at night and three in the morning. So I put the three in here and the one at night and I just rubber banded these together. That's how it's foolproof. You don't have to pack three different bottles and remember what combination.* [A14]

Community members can participate in action work, often providing support for improved illness management —such as taking medication consistently, keeping doctor's appointments, or healthy eating. A12 explained how she kept her child's school updated after changes in treatment plans:

> *I do a separate inhaler for school, I have a current prescription, I have Dr. A specifically sign on the paper saying this is the plan, this is how much she gets it if she needs it, she can or cannot carry it with her.* [A12]

People also update their tracking routine to sustain engagement. One person described his motivation for trying new tools and methods:

> *If you find a way to evolve the process frequently enough and meaningfully enough that you're still excited about it as you go on, then I think that's really powerful.* [Q3]

Dependencies and Workflow

Li et al [11] and Epstein et al [16] describe similar sequences of tracking work, with starting and ending points. Swan [42] describes types of work but no sequence, start, or end. We did not find a specific sequence but did find dependencies. We did not identify a definitive starting point but did identify common situations that trigger tracking. Owing to the ongoing nature of chronic illness, we did not observe a final ending to tracking, although 1 QS speaker did talk about tracking less frequently or discontinuing tracking during periods of better symptom control (Q14). Table 3 describes the dependencies among types of work. Table 4 describes 2 common situations that trigger tracking.

We found evidence for bidirectional transitions between each type of work, often involving handoffs between actors, and found that work does often overlap, corroborating Epstein et al [16] and Costa Figueirido et al [14]. For example, a parent may receive communication from a child's teacher concerning a new symptom. Examining this exchange reveals a chain of rapid and interleaved tasks:

collection (teacher: symptom) → *integration* (teacher: past symptom data) → *reflection* (teacher: is this a new symptom?) → *communication* (teacher→parent: possible new symptom) → *integration* (parent: past symptom data) → *reflection* (parent: this is a new symptom) → *information/communication/ integration/ reflection* (parent: new symptom)

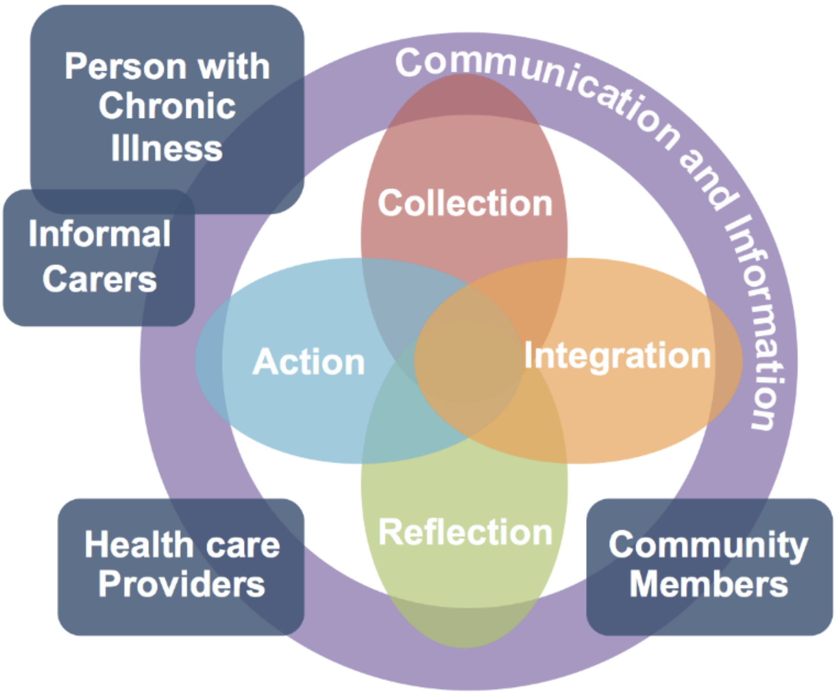
Table 3. Dependencies among types of work.

Dependency	Description
(Collection, Information) → Reflection	Reflection work cannot take place without some kind of collection or information work (eg, a weight measurement or list of medication side effects)
Collection → Integration	Integration work cannot take place without at least 1 data point each of 2 types of data (eg, a meal photo with a blood glucose measurement)
(Information, Communication, Reflection) → Action	As we defined it, action work is a change of plan for managing the illness and must be based on the outcome of other work, usually information, communication, or reflection work

Table 4. Common triggers for tracking.

Trigger	Description
Collection	Collection (ie, observed symptoms or abnormal test result) leading to reflection (ie, “what does that high blood pressure result mean?”)
Information or Communication	Incidental information or communication (ie, reading a magazine article or talking to a friend) gives rise to reflection (ie, “is that why I’ve been feeling tired?”) and collection (ie, “let’s investigate”)

Figure 1. The Conceptual Model of Shared Health Informatics (CoMSHI) showing the work and social context of tracking in support of chronic illness management and the interplay between components. Actors are the person with chronic illness, informal carers, community members, and health care providers. The work in which those actors engage includes communication, information, collection, integration, reflection, and action. Work is done in no particular order, and types of work can overlap. All actors may engage in work and interact with each other around that work.



The parent’s work in the last step combines information work interleaved with integration, reflection, and communications work concerning how the new symptom and information fit with her previous understanding of information and data (eg, reflection-in-action [54]).

We developed a conceptual model of the tracking process that supports chronic illness management based on what we learned.

Conceptual Model of Shared Health Informatics

To bridge the gap between current informatics models and important characteristics of tracking for chronic illness management, we proposed the CoMSHI (pronounced *com-she*; Figure 1). The CoMSHI is based on insights from previous research (eg, [11,14,20,21,24,42,47]) and new data analysis about tracking behavior and social interactions. It portrays the actors and work, described above, that drive successful tracking in support of chronic illness management. Actors perform work in no particular order, and work can be ongoing and overlapping. All actors may engage in work and communicate around that work.

Discussion

In this study, we aimed to construct a model in response to informatics literature indicating a need for better representation of the unique challenges and context around tracking to support illness management [13,14,47]. We combined insights from the literature with an analysis of 69 interview and video transcripts to develop the CoMSHI.

Contributions

The CoMSHI extends previous work on personal and health informatics models; Table 5 summarizes a comparison with that previous work. Our model is unique in describing the relationships among people and work involved in tracking in support of chronic illness management and emphasizes communication and shared work.

Table 6 summarizes the contributions of this study. A valuable extension to previous personal and health informatics models is the inclusion of *carers*. Although Pew's health tracking survey [6] found that 12% of trackers track for someone else, no other informatics model includes carers as primary actors and trackers. To address this gap observed in the literature and our data analysis, we have highlighted that carers often assume a critical role in tracking to manage chronic illness. We modified Li et al's stages of work to unconstrained transitions between work because people managing chronic illness do not progress through a sequence of stages but continuously and iteratively work in support of their health [7]. We redefined the *preparation* stage to *information* work. This reflects the ongoing knowledge building that supports the other work and actors. Our analysis also showed that shared work among the person with chronic illness, carer, community members, and health care providers was key to successful tracking, in line with findings of other research outlined in the related literature.

Table 5. Characteristics of the Conceptual Model of Shared Health Informatics (CoMSHI) compared with models from studies by Li et al, Epstein et al, Swan, and Murnane et al.

Model	Model description and basis	Role of tracker	Work	Roles of others	Outcomes
Stage-Based Model of Personal Informatics [11]	Literature analysis, empirical study defining personal informatics	One person who performs all work	Preparation, collection, integration, reflection, action	N/A[a]	Increased self-knowledge, informed action
Lived Informatics Model of Personal Informatics [16]	Literature analysis, empirical study defining lived informatics	One person who performs all work	Deciding, Selecting, tracking and acting, lapsing	N/A	Increased self-knowledge, informed action, lapsed tool use with possible resumption
Patient-Driven Health Care Model [42]	Description of patient-driven health care	One patient who performs all work	Research, treat, intervene, experiment, track, measure	Patient initiates contact with peers and professionals	Self-expression, enhancement, prevention, cure, normalization, improvement
Model of the Sociotechnical Ecology Surrounding Serious Mental Illness Management [43]	Literature analysis, empirical study defining social relations in managing severe mental illness	One patient who performs work and is influenced by external actors and contexts	N/A	Patient interacts with close ties, institutions, sociocultural context	Interpersonal comparisons and baselines, mitigation and management of crises
Conceptual Model of Shared Health Informatics	Literature analysis, empirical study defining people and work in tracking to manage chronic illness	One or more people who communicate and share tracking work	Communication, information, collection, integration, reflection, action	Part of the social ecology communicating and supporting tracking work	Increased knowledge, communication, informed action

[a] N/A: not applicable.

Table 6. Contributions of the Conceptual Model of Shared Health Informatics (CoMSHI).

Contribution	Description
Carer as primary actor	Carers often assume a critical role in tracking to manage chronic illness
Communication work	Communication work supports interactions among actors around tracking work
Information and communication work support tracking practice	Information and communication work are the backbone enabling exchange of ideas and insights as well as transitions between work
Distributed work	Tracking work is distributed across multiple actors rather than resting only with one person
No prescribed work sequence	Work is ongoing, nonsequential, and sometimes overlapping rather than linear and time-limited

Implications for Design

As are previous models, the CoMSHI is agnostic to specific tools used or data elements collected. Rather, the model describes the relationships among work and people that health informatics tools need to accommodate. New health informatics tools would better align with the experience of people involved in tracking for chronic illness management if designed to support both the types of work and actors involved, thus promoting effective management and potentially improving health outcomes. On the basis of the CoMSHI, we recommend that, early in the design process, designers determine the extent of the tracking practice their tool will support and then define the functionality necessary to facilitate shared work and transitions between people and types of work. Any one tool does not need to support all aspects of tracking work, but designers must critically consider how to empower people to track the data they need, collaborate with whom they choose, and transition between tools that support other tracking tasks. The first step in accomplishing this goal is to develop a deep understanding of the users of the technology, their goals for tracking, and their illness or illnesses. Without thoughtful engagement with people and work, designers will find it difficult to create truly usable and useful technology for those they serve.

Limitations and Future Work

This model is based upon transcript analysis of people managing chronic illnesses and as such we can only claim that it applies to that context. However, based on our understanding of the literature, it may apply to tracking in other health contexts such as for people who are hospitalized [47] or people with cancer [52,55]. We also did not interview representative health care providers or community members for their perspectives. Further research is needed to evaluate generalizability. Also, several articles [28,29,56] assert that the privacy policies around health information technology, especially patient portals, are insufficient to effectively support the needs of people with chronic illness and carers. Our study suggests that further research should also consider the role of community members to ascertain how to best support the work and social ecosystem of tracking in support of chronic illness.

Conclusions

For people managing chronic illness, effective tracking improves health outcomes. Health informatics tools intend to help but they often fall short of supporting the true range of work and people involved. Furthermore, current research and tools often focus on *personal* informatics, *self*-management, or *self*-tracking—limiting how we think about and design to support tracking for chronic illness management. Understanding the shared work of tracking can inform the design of systems to support the reality of managing chronic illness [7,8]. One QS speaker asserts: “in chronic diseases, health is not created in healthcare (Q16),” emphasizing that she cannot rely solely on health care providers and her tracking practice supports her health in the life outside the clinic.

This study has contributed a model of the work and social context of tracking in support of chronic illness management to advance the understanding of how to support successful health tracking. The CoMSHI gives insight into the processes used by people who successfully manage a chronic illness as well as the context in which they work. The CoMSHI expands on its predecessors by (1) including informal carers, (2) emphasizing the shared nature of tracking work, and (3) characterizing work as ongoing and nonsequential. This new model demonstrates the fluidity of the tracking process and situates the work of tracking in its social context. Most importantly, this work underscores the impossibility of isolating tracking work from the social environment of people managing chronic illness, and designers must consider the shared aspects of tracking when designing health informatics tools. Although previous models focus on a single person engaging in tracking work, the CoMSHI emphasizes that it is only part of the puzzle. As one speaker expressed, “It’s not just technology, it’s people” (Q21).

Acknowledgments

The authors would like to thank the people managing chronic illness whose experiences formed the basis for this research. They also thank Kim Unertl for her insights into defining the model and the reviewers for their constructive criticism. This project was supported by the Agency for Healthcare Research and Quality (AHRQ; grant #R01HS021590), the National Institutes of Health (NIH) National Library of Medicine Biomedical and Health Informatics Training Grant at the University of Washington (grant #T15LM007442), the National Science Foundation (grant #IIS-1553167), and the National Center for Advancing Translational Sciences Grant at the University of North Carolina at Chapel Hill (grant #UL1TR002489). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ or NIH.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2008 Jun 1. Noncommunicable diseases URL: https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2019-01-31] [WebCite Cache ID 75qJzFLqo]
2. Anderson G. Robert Wood Johnson Foundation. 2010 Jan 01. Chronic Care: Making the Case for Ongoing Care URL: https://www.rwjf.org/en/library/research/2010/01/chronic-care.html [accessed 2019-01-31] [WebCite Cache ID 6ybUoqNwm]

3. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. Prev Chronic Dis 2014 Apr 17;11:E62 [FREE Full text] [doi: 10.5888/pcd11.130389] [Medline: 24742395]
4. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
5. Centers for Disease Control and Prevention. 2016. Diabetes at a Glance: Working to Reverse the US Epidemic URL: https://www.cdc.gov/diabetes/library/factsheets.html [accessed 2019-01-31] [WebCite Cache ID 75qKRlV6R]
6. Fox S, Duggan M. Pew Research Center. 2013 Jan 28. Tracking for Health URL: http://www.pewinternet.org/2013/01/28/tracking-for-health [accessed 2019-01-31] [WebCite Cache ID 6ybUvrzhH]
7. Unruh KT, Pratt W. The invisible work of being a patient and implications for health care:"[the doctor is] my business partner in the most important business in my life, staying alive". Conf Proc Ethnogr Prax Ind Conf 2008 Nov;2008(1):40-50 [FREE Full text] [doi: 10.1111/j.1559-8918.2008.tb00093.x] [Medline: 21709742]
8. 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]
9. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2016 Presented at: CHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Diego, CA p. 1109-1113 URL: http://europepmc.org/abstract/MED/28503678 [doi: 10.1145/2858036.2858045]
10. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collecting and exploring personal data. USA: ACM; 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26-May 01, 2014; Toronto, CA p. 1143-1152. [doi: 10.1145/2556288.2557372]
11. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA. [doi: 10.1145/1753326.1753409]
12. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
13. Valdez RS, Brennan PF. Exploring patients' health information communication practices with social network members as a foundation for consumer health IT design. Int J Med Inform 2015 May;84(5):363-374. [doi: 10.1016/j.ijmedinf.2015.01.014] [Medline: 25704761]
14. Figueiredo MC, Caldeira C, Reynolds TL, Victory S, Zheng K, Chen Y. Self-tracking for fertility care: collaborative support for a highly personalized problem. In: Proceedings of the ACM on Human-Computer Interaction. New York, NY, USA: ACM; 2017 Presented at: ACM on Human-Computer Interaction; November 2017; Portland, Oregon, USA, p. 1-21. [doi: 10.1145/3134671]
15. Quantified Self. URL: http://quantifiedself.com/ [accessed 2019-01-31] [WebCite Cache ID 6ybUb4ZkX]
16. Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY, USA: ACM; 2015 Presented at: International Joint Conference on Pervasive and Ubiquitous Computing; September 07-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2804250]
17. Andrew A, Borriello G, Fogarty J. Simplifying mobile phone food diaries: design and evaluation of a food index-based nutrition diary. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. 2013 May 23 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; May 5-8, 2013; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252101]
18. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss 'N' turn: smartphone as sleep and sleep quality detector. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 01, 2014; Toronto, Ontario, Canada. [doi: 10.1145/2556288.2557220]
19. Purpura S, Schwanda V, Williams K, Stubler W, Sengers P. Fit4Life: the design of a persuasive technology promoting healthy behavior and ideal weight. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2011; Vancouver, BC, Canada p. 423-432. [doi: 10.1145/1978942.1979003]
20. Epstein DA, Jacobson BH, Bales E, McDonald DW, Munson SA. From. New York, NY, USA: ACM; 2015 Presented at: "Nobody Cares" to "Way to Go!": A Design Framework for Social Sharing in Personal Informatics. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, BC, Canada p. 1622-1636. [doi: 10.1145/2675133.2675135]
21. Chung CF, Agapie E, Schroeder J, Mishra S, Fogarty J, Munson SA. When personal tracking becomes social: examining the use of instagram for healthy eating. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing System. New York, NY, USA: ACM; 2017 Presented at: CHI Conference on Human Factors in Computing System; May 06-11, 2017; Denver, CO p. 1674-1687 URL: http://europepmc.org/abstract/MED/28516174 [doi: 10.1145/3025453.3025747]

22. Eastwood CA, Travis L, Morgenstern TT, Donaho EK. Weight and symptom diary for self-monitoring in heart failure clinic patients. J Cardiovasc Nurs 2007;22(5):382-389. [doi: 10.1097/01.JCN.0000287027.49628.a7] [Medline: 17724420]
23. Baos V, Ester F, Castellanos A, Nocea G, Caloto MT, Gerth WC, I-Max Study Group. Use of a structured migraine diary improves patient and physician communication about migraine disability and treatment outcomes. Int J Clin Pract 2005 Mar;59(3):281-286. [doi: 10.1111/j.1742-1241.2005.00469.x] [Medline: 15857323]
24. MacLeod H, Tang A, Carpendale S. Personal informatics in chronic illness management. In: Proceedings of Graphics Interface 2013. Toronto, ON, Canada: Canadian Information Processing Society; 2013 Presented at: Graphics Interface 2013; May 29-31, 2013; Regina, SK, Canada.
25. Chung CF, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, NY, USA: ACM; 2016 Presented at: Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 02, 2016; San Fransisco, CA p. 770-786 URL: http://europepmc.org/abstract/MED/28516171 [doi: 10.1145/2818048.2819926]
26. Mamykina L, Heitkemper EM, Smaldone AM, Kukafka R, Cole-Lewis HJ, Davidson PG, et al. Personal discovery in diabetes self-management: discovering cause and effect using self-monitoring data. J Biomed Inform 2017 Dec;76:1-8. [doi: 10.1016/j.jbi.2017.09.013] [Medline: 28974460]
27. Pina LR, Sien SW, Ward T, Yip JC, Munson SA, Fogarty J, et al. From personal informatics to family informatics: understanding family practices around health monitoring. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, NY, USA: ACM; 2017 Presented at: Conference on Computer Supported Cooperative Work and Social Computing; February 25-March 01, 2017; Portland, OR. [doi: 10.1145/2998181.2998362]
28. Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller DP, Smith D, et al. Insights into older adult patient concerns around the caregiver proxy portal use: qualitative interview study. J Med Internet Res 2018 Nov 02;20(11):e10524 [FREE Full text] [doi: 10.2196/10524] [Medline: 30389654]
29. Sarkar U, Bates DW. Care partners and online patient portals. J Am Med Assoc 2014 Jan;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
30. Gorm N, Shklovski I. Steps, choices and moral accounting: observations from a step-counting campaign in the workplace. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, NY, USA: ACM; 2016 Presented at: Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 02, 2016; San Fransisco, CA p. 148-159. [doi: 10.1145/2818048.2819944]
31. Paton C, Hansen M, Fernandez-Luque L, Lau AY. Self-tracking, social media and personal health records for patient empowered self-care. Contribution of the IMIA social media working Group. Yearb Med Inform 2012;7:16-24. [Medline: 22890336]
32. Schroeder J, Hoffswell J, Chung CF, Fogarty J, Munson S, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, NY, USA: ACM; 2017 Presented at: Conference on Computer Supported Cooperative Work and Social Computing; February 25-March 01, 2017; Portland, OR p. 1726-1739 URL: http://europepmc.org/abstract/MED/28516172 [doi: 10.1145/2998181.2998276]
33. Cordeiro F, Epstein DA, Thomaz E, Bales E, Jagannathan AK, Abowd GD, et al. Barriers and negative nudges: exploring challenges in food journaling. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2015 Presented at: Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea p. 1159-1162 URL: http://europepmc.org/abstract/MED/26894233 [doi: 10.1145/2702123.2702155]
34. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for people With rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
35. Choi YK, Demiris G, Lin SY, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
36. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
37. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
38. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]

39. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
40. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
41. Mamykina L, Mynatt E, Davidson P, Greenblatt D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 05-10, 2008; Florence, Italy p. 477-486. [doi: 10.1145/1357054.1357131]
42. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
43. Murnane EL, Walker TG, Tench B, Voida S, Snyder J. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. New York, NY, USA: ACM; 2018 Nov 01 Presented at: Proceedings of the ACM on Human-Computer Interaction - CSCW; November 2018; New York p. 1-27. [doi: 10.1145/3274396]
44. Whooley M, Ploderer B, Gray K. On the integration of self-tracking data amongst quantified self members. In: Proceedings of the 28th International BCS Human Computer Interaction Conference. 2014 Presented at: 28th International BCS Human Computer Interaction Conference; September 9-12, 2014; Southport, UK. [doi: 10.14236/ewic/hci2014.16]
45. Rooksby J, Rost M, Morrison A, Chalmers M. Personal tracking as lived informatics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 01, 2014; Toronto, ON, Canada p. 1163-1172. [doi: 10.1145/2556288.2557039]
46. Bronfenbrenner U. Ecological systems theory (1992). In: Making Human Beings Human: Bioecological Perspectives on Human Development. Thousand Oaks, CA: Sage Publications; 2005.
47. Mishra SR, Miller AD, Haldar S, Khelifi M, Eschler J, Elera RG, et al. Supporting collaborative health tracking in the hospital: patients' perspectives. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada URL: http://europepmc.org/abstract/MED/29721554 [doi: 10.1145/3173574.3174224]
48. Holtzblatt K, Beyer H. Contextual Design: Design for Life. 2nd Edition. Amsterdam: Morgan Kaufmann; Dec 14, 2017.
49. Wagner EH. The role of patient care teams in chronic disease management. Br Med J 2000 Feb 26;320(7234):569-572 [FREE Full text] [Medline: 10688568]
50. Grande SW, Sherman LD. Too important to ignore: leveraging digital technology to improve chronic illness management among black men. J Med Internet Res 2018 May 14;20(5):e182 [FREE Full text] [doi: 10.2196/jmir.9434] [Medline: 29759956]
51. Veinot TC, Meadowbrooke CC, Loveluck J, Hickok A, Bauermeister JA. How "community" matters for how people interact with information: mixed methods study of young men who have sex with other men. . J Med Internet Res 2013 Feb 21;15(2):e33 [FREE Full text] [doi: 10.2196/jmir.2370] [Medline: 23428825]
52. Weiss JB, Lorenzi NM. Online communication and support for cancer patients: a relationship-centric design framework. AMIA Annu Symp Proc 2005:799-803 [FREE Full text] [Medline: 16779150]
53. Skeels MM, Unruh KT, Powell C, Pratt W. Catalyzing social support for breast cancer patients. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA p. 173-182 URL: http://europepmc.org/abstract/MED/21654894 [doi: 10.1145/1753326.1753353]
54. Schon DA. The Reflective Practitioner: How Professionals Think In Action. New York, NY: Basic Books; 1984.
55. Patel RA, Klasnja P, Hartzler A, Unruh KT, Pratt W. Probing the benefits of real-time tracking during cancer care. AMIA Annu Symp Proc 2012;2012:1340-1349 [FREE Full text] [Medline: 23304413]
56. Zulman DM, Piette JD, Jenchura EC, Asch SM, Rosland AM. Facilitating out-of-home caregiving through health information technology: survey of informal caregivers' current practices, interests, and perceived barriers. J Med Internet Res 2013 Jul 10;15(7):e123 [FREE Full text] [doi: 10.2196/jmir.2472] [Medline: 23841987]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CoMSHI: Conceptual Model of Shared Health Informatics
HIPAA: Health Insurance Portability and Accountability Act
NIH: National Institutes of Health
QS: Quantified Self

Edited by G Eysenbach; submitted 19.04.18; peer-reviewed by R Valdez, L Novak; comments to author 25.08.18; revised version received 31.01.19; accepted 18.02.19; published 29.04.19.

Please cite as:
Vizer LM, Eschler J, Koo BM, Ralston J, Pratt W, Munson S
"It's Not Just Technology, It's People": Constructing a Conceptual Model of Shared Health Informatics for Tracking in Chronic Illness Management
J Med Internet Res 2019;21(4):e10830
URL: http://www.jmir.org/2019/4/e10830/
doi:10.2196/10830
PMID:31033452

Original Paper

Crowdsourcing for Food Purchase Receipt Annotation via Amazon Mechanical Turk: A Feasibility Study

Wenhua Lu[1], PhD; Alexandra Guttentag[2], BA; Brian Elbel[3,4], PhD, MPH; Kamila Kiszko[3], MPH; Courtney Abrams[3], MA; Thomas R Kirchner[2], PhD

[1]Department of Childhood Studies, Rutgers, The State University of New Jersey, Camden, NJ, United States
[2]College of Global Public Health, New York University, New York, NY, United States
[3]School of Medicine, New York University, New York, NY, United States
[4]Robert F Wagner Graduate School of Public Service, New York University, New York, NY, United States

Corresponding Author:
Wenhua Lu, PhD
Department of Childhood Studies
Rutgers, The State University of New Jersey
329, Cooper Street
Camden, NJ, 08102
United States
Phone: 1 856 225 6083
Email: w.lu@rutgers.edu

Abstract

Background: The decisions that individuals make about the food and beverage products they purchase and consume directly influence their energy intake and dietary quality and may lead to excess weight gain and obesity. However, gathering and interpreting data on food and beverage purchase patterns can be difficult. Leveraging novel sources of data on food and beverage purchase behavior can provide us with a more objective understanding of food consumption behaviors.

Objective: Food and beverage purchase receipts often include time-stamped location information, which, when associated with product purchase details, can provide a useful behavioral measurement tool. The purpose of this study was to assess the feasibility, reliability, and validity of processing data from fast-food restaurant receipts using crowdsourcing via Amazon Mechanical Turk (MTurk).

Methods: Between 2013 and 2014, receipts (N=12,165) from consumer purchases were collected at 60 different locations of five fast-food restaurant chains in New Jersey and New York City, USA (ie, Burger King, KFC, McDonald's, Subway, and Wendy's). Data containing the restaurant name, location, receipt ID, food items purchased, price, and other information were manually entered into an MS Access database and checked for accuracy by a second reviewer; this was considered the *gold standard*. To assess the feasibility of coding receipt data via MTurk, a prototype set of receipts (N=196) was selected. For each receipt, 5 turkers were asked to (1) identify the receipt identifier and the name of the restaurant and (2) indicate whether a beverage was listed in the receipt; if yes, they were to categorize the beverage as cold (eg, soda or energy drink) or hot (eg, coffee or tea). Interturker agreement for specific questions (eg, restaurant name and beverage inclusion) and agreement between turker consensus responses and the gold standard values in the manually entered dataset were calculated.

Results: Among the 196 receipts completed by turkers, the interturker agreement was 100% (196/196) for restaurant names (eg, Burger King, McDonald's, and Subway), 98.5% (193/196) for beverage inclusion (ie, hot, cold, or none), 92.3% (181/196) for types of hot beverage (eg, hot coffee or hot tea), and 87.2% (171/196) for types of cold beverage (eg, Coke or bottled water). When compared with the gold standard data, the agreement level was 100% (196/196) for restaurant name, 99.5% (195/196) for beverage inclusion, and 99.5% (195/196) for beverage types.

Conclusions: Our findings indicated high interrater agreement for questions across difficulty levels (eg, single- vs binary- vs multiple-choice items). Compared with traditional methods for coding receipt data, MTurk can produce excellent-quality data in a lower-cost, more time-efficient manner.

(J Med Internet Res 2019;21(4):e12047) doi:10.2196/12047

KEYWORDS

Amazon Mechanical Turk; food purchase receipt; crowdsourcing; feasibility; reliability; validity

Introduction

The decisions that individuals make about the food and beverage products they purchase and consume directly influence their energy intake and dietary quality and may lead to excess weight gain and obesity [1-3]. Research supports the notion that decision making related to food consumption may act as a potential mediator between the neighborhood food environment and individual dietary intake [4-7], but assessment of dietary behavior can be problematic [1]. Leveraging new sources of data on food and beverage purchase behavior, therefore, could provide novel insights into food and beverage decision making.

Food purchase receipts contain information about all foods and beverages purchased by individuals and households from different sources, such as fast-food restaurants, grocery stores, and convenience or corner stores [1]. Compared with retrospective self-reports, receipts can contribute more objective data, thereby avoiding social desirability influence and recall bias [8]. Unfortunately, accurately and reliably annotating large numbers of receipts and images has been a logistical bottleneck inhibiting their widespread use. Typically, academic researchers depend on undergraduate and graduate research assistants to extract data; research progress then depends on the ebb and flow of the semester. Further, each receipt must be carefully reviewed, which takes several minutes. As a result, it can take weeks, months, or even years to process receipt data, especially when large datasets are being handled and/or subjective reasoning is needed.

In the past decade, crowdsourcing has become increasingly popular due to its time-saving and cost-effective qualities [9]. In crowdsourcing, potentially large jobs are broken into many microtasks that are then outsourced directly to individual workers via public solicitation [10]. As the leading and most well-established online crowdsourcing service, Amazon Mechanical Turk (MTurk) enables researchers and businesses, identified as requesters, to recruit anonymous online workers (ie, turkers) worldwide to complete Human Intelligence Tasks (HITs) (ie, tasks that cannot be entirely automated and require human intelligence) at relatively low cost [11]. MTurk offers a basic user interface for simple tasks and a powerful application programming interface for developers to build a platform that uses their services [10,11].

Since its inception, MTurk has been used primarily by researchers in nonmedical fields (eg, psychology, marketing, management, business, political science, computer science, and neuroscience) to do data processing, including data extraction, transcription, translation, and sentiment analysis [12-18]. Emerging studies in recent years have also applied MTurk in various disciplines of health [12-14]. For example, a group of researchers pioneered the use of crowdsourcing technology in public health research and utilized a custom MTurk interface for analyzing mobile phone photographs of retail point-of-sale tobacco marketing [19,20]. Over the course of one typical implementation, 299 turkers completed more than 23,000 tasks at a total cost of US $2500 in less than 24 hours. Results of the crowdsourced photo-only assessments had an excellent level of correspondence to the traditional field survey data, which demonstrated the tremendous potential and reliability of MTurk as a medium for analyzing health-related data in a low-cost, time-efficient way [19,20].

Despite its growing popularity, MTurk has not yet been used to annotate data from food and beverage purchase receipts. This study, therefore, takes an initial foray into assessing the feasibility, reliability, and validity of processing fast-food restaurant receipt data using MTurk.

Methods

Overview

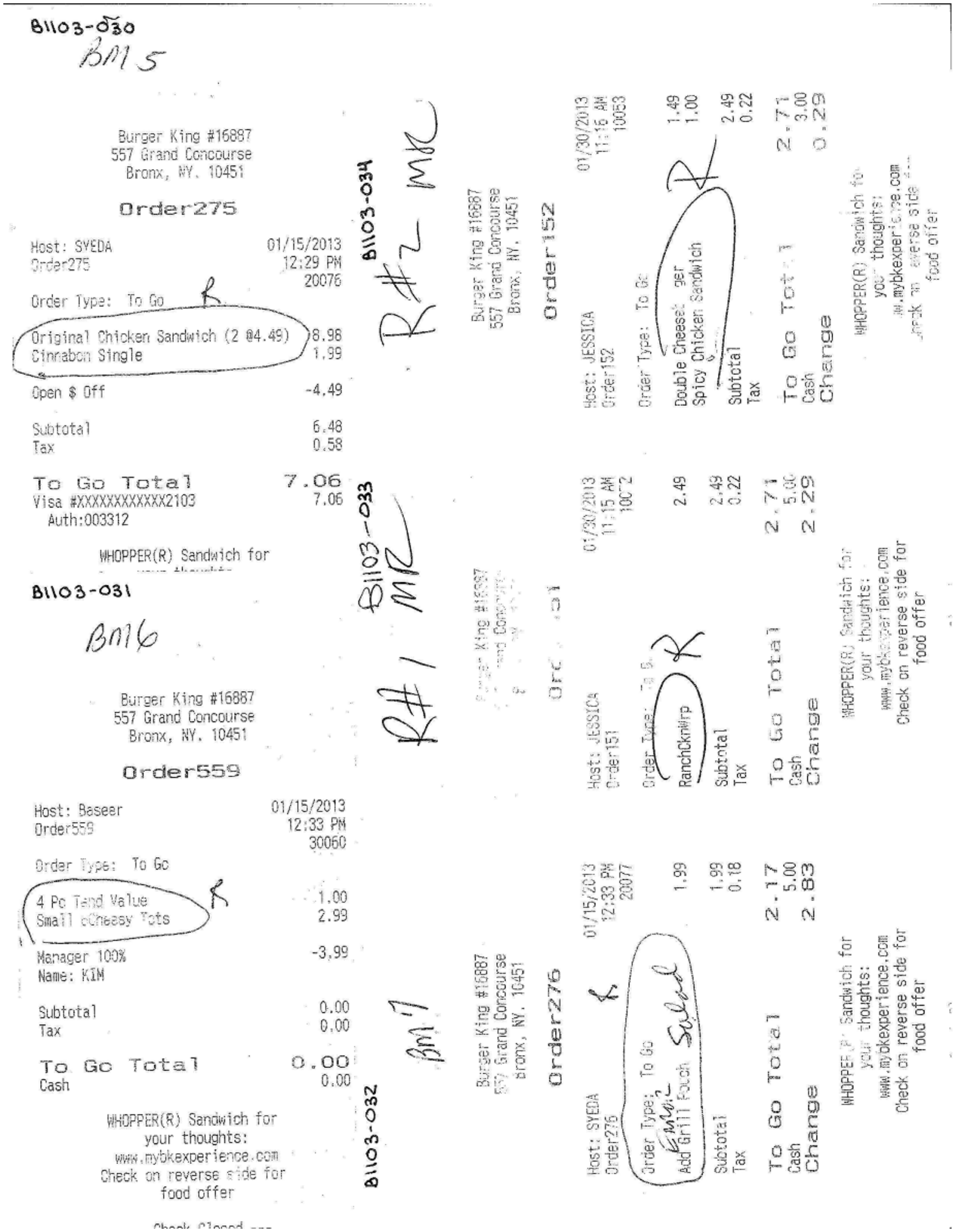
Our study consisted of three phases: First, data from a large number of food and beverage purchase receipts were obtained through a traditional in-laboratory, manual data extraction method and were confirmed for accuracy to serve as the *gold standard*; second, an MTurk project was set up and a group of turkers were recruited to extract some prespecified required data from a representative sample of the receipts; and third, the data processed through MTurk were compared with the gold standard and evaluated for reliability and validity. Details of each step are described in the following sections.

Step 1: Data Collection and Manual Data Extraction

Between August 2013 and May 2014, receipts were collected from consumer purchases at 60 different locations of five fast-food restaurant chains in New Jersey and New York City, USA: Burger King, KFC, McDonald’s, Subway, and Wendy’s. Data collectors stood outside of the restaurants and asked entering customers to save their receipts; upon leaving, they were asked to hand over their receipt and identify which items on the receipt they purchased for their own consumption. Altogether, three rounds of data collection were conducted within the project period, and a total of 12,165 receipts were collected. Detailed data collection procedure, the number of locations surveyed, and receipts collected by location and restaurant chain have been described previously [21].

After each round of receipt collection, research assistants pasted receipts on single printer paper sheets next to each other—about 4-6 per page—and scanned them into a database (see Figure 1). Unusual receipts were flagged and reported in the database; for example, those where details of exactly what was purchased by the customers were missing (ie, the receipt was not clearly marked, the ink rubbed off, or the receipt was not itemized).

Figure 1. Sample food purchase receipts.



B1103-030
BM 5

Burger King #16887
557 Grand Concourse
Bronx, NY. 10451

Order275

Host: SYEDA 01/15/2013
Order275 12:29 PM
20076

Order Type: To Go R

Original Chicken Sandwich (2 @4.49)	8.98
Cinnabon Single	1.99
Open $ Off	-4.49
Subtotal	6.48
Tax	0.58
To Go Total	7.06
Visa #XXXXXXXXXXXX2103	7.06
Auth:003312	

WHOPPER(R) Sandwich for

B1103-034
R#2 MR

B1103-031
BM6

Burger King #16887
557 Grand Concourse
Bronx, NY. 10451

Order559

Host: Baseer 01/15/2013
Order559 12:33 PM
30060

Order Type: To Go

4 Pc Tend Value	1.00
Small Cheesy Tots	2.99
Manager 100%	-3.99
Name: KIM	
Subtotal	0.00
Tax	0.00
To Go Total	0.00
Cash	0.00

R

WHOPPER(R) Sandwich for
your thoughts:
www.mybkexperience.com
Check on reverse side for
food offer

B1103-033
MR
R#1

BM7
B1103-032

Burger King #16887
557 Grand Concourse
Bronx, NY. 10451

Order152

Host: JESSICA 01/30/2013
Order152 11:16 AM
10053

Order Type: To Go

Double Cheeseburger	1.49
Spicy Chicken Sandwich	1.00
Subtotal	2.49
Tax	0.22
To Go Total	2.71
Cash	3.00
Change	0.29

R

WHOPPER(R) Sandwich for
your thoughts:
www.mybkexperience.com
Check on reverse side for
food offer

Host: JESSICA 01/30/2013
Order151 11:15 AM

Order Type: To Go

RanchCknWrp	2.49
Subtotal	2.49
Tax	0.22
To Go Total	2.71
Cash	5.00
Change	2.29

R

WHOPPER(R) Sandwich for
your thoughts:
www.mybkexperience.com
Check on reverse side for
food offer

Burger King #16887
557 Grand Concourse
Bronx, NY. 10451

Order276

Host: SYEDA 01/15/2013
Order276 12:33 PM
20077

Order Type: To Go R

Add Grill Pouch	1.99
Subtotal	1.99
Tax	0.18
To Go Total	2.17
Cash	5.00
Change	2.83

Salad

WHOPPER(R) Sandwich for
your thoughts:
www.mybkexperience.com
Check on reverse side for
food offer

Following that, the data containing the receipt identifier, restaurant name, food items purchased, purchase date, total price, etc, were extracted from individual receipts; manually entered into an MS Access database by a research assistant; and checked for accuracy by a second research assistant 1-10 months following the initial data entry. The receipt data obtained through such a traditional in-laboratory manual data extraction method served as the gold standard in this study.

Step 2: Setting Up the Amazon Mechanical Turk Task and Crowdsourcing Workflow

To assess the feasibility of MTurk as an alternative tool for processing receipt data, some of the data extracted manually above was recollected using crowdsourcing via MTurk.

Specifically, two separate tasks were set up on MTurk. The first was an *expert* task, in which turkers were asked to crop one receipt per page at a time from the original pages with 4-6 receipts per page. An *expert* task in MTurk means that requesters trust one turker to do the assignment, rather than having multiple turkers do it and agree on an answer; such tasks are usually simple and do not involve extensive human reasoning and interpretation. First, scanned PDF documents (8.5 x 11 inches) with multiple receipts in different orientations on each page were uploaded onto MTurk. Following that, the MTurk Expert HIT was launched and turkers were recruited to crop each individual receipt using a crop tool and orient the receipts in a readable fashion. An instructional video was included to guide turkers in using the software correctly. This task was completed preceding this study for all receipts (N=12,165), with one receipt on one page.

The next MTurk HIT was a *consensus* task—the focus of this study—which required multiple turkers working on the same assignment and then checking the agreement among their responses. For this task, entitled “Receipt Information,” turkers were requested to identify required information from the food purchase receipts and respond to a series of questions based on the information they identified. As illustrated in Figure 2, a brief description was included under the title of the project: “Please gather the following information related to food purchase from a receipt.” For each receipt, turkers were asked to answer questions based on the following four tasks: (1) write down the receipt identifier, (2) choose the name of the restaurant from a drop-down list, (3) indicate whether a beverage was listed in the receipt, and (4) if a beverage was listed, categorize the beverage as cold beverage or hot beverage.

Specifically, question 1 required textual responses; for each individual receipt, turkers were requested to type in a unique identifier composed of letters and numbers (eg, B1103-036 and S2109A-022). Questions 2-4 included multiple-choice questions, which required subjective judgment at different difficulty levels (ie, single- vs binary-choice items). Considering that most information on food purchase receipts can be obtained through either textual responses or multiple-choice questions, it is reasonable to assume that if a turker can understand and respond accurately to these four exemplary questions, he or she could identify other data from food purchase receipts as well. For demonstration purposes, therefore, instead of using all of the 12,165 receipts, a prototypical sample of receipts (N=196) were used for this study, all of which were clearly marked with zero or only one beverage item on each receipt.

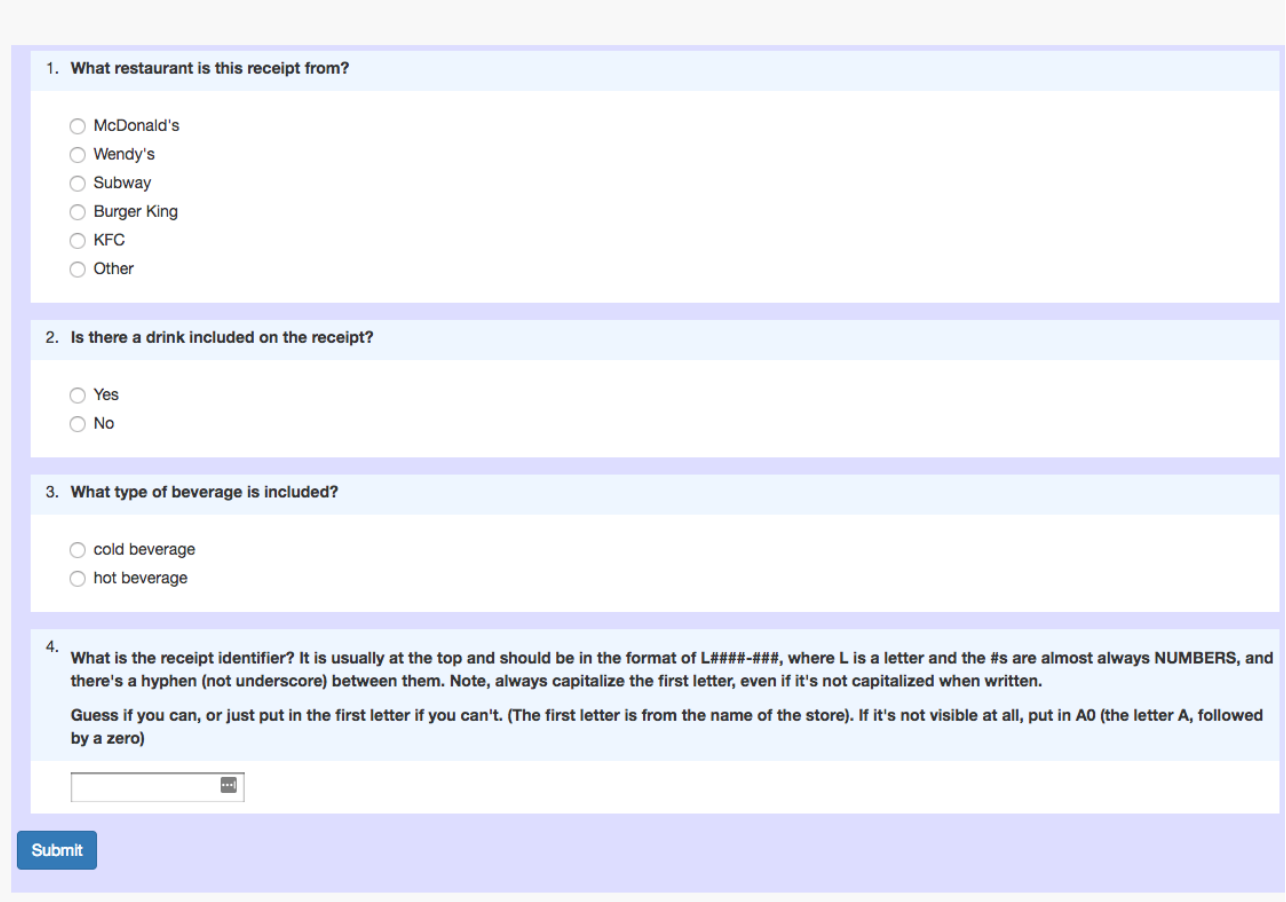
After the prototypical receipts were selected and the MTurk HIT was set up, we started to invite turkers to work on the consensus task. To avoid spammers and control the quality of turkers, we screened the turkers by setting the minimum prior approval rating to 99%, meaning that at least 99% of a turker’s answers to the MTurk HITs that they have completed to date were approved by the requester. Turkers’ locations were further restricted to the United States, as previous studies have suggested that language, cultural background, and ethnicity can significantly influence people’s comprehension of culture-related information such as food choices [16,22,23]. Turkers were paid US $0.06 for interpreting each receipt, which was anticipated to take 60-90 seconds. Once a turker began processing a receipt, he or she had a maximum of 5 minutes to complete it. A turker could analyze as many receipts as he or she wanted.

One critical question in using crowdsourcing via MTurk is setting the minimum and maximum number of turkers who will complete each assignment (ie, the number of repetitions that each receipt will receive until consensus is achieved for each question in the study). Intuitively, 2 would be the absolute minimum in order to reach an agreement on responses to a question. However, setting 2 as the minimum number of repetitions can incur an incorrect agreed-upon answer if both turkers provide the same incorrect responses to a question. A minimum of 3 creates a majority, but the question agreement threshold (QAT) of 67% (2/3) is insufficiently low and an incorrect consensus can still be reached if 2 of the 3 turkers agree on an incorrect answer. A minimum of 4 is acceptable with an interturker agreement of 75% (3/4), but a consensus cannot be achieved if 2 turkers agree on one response while the other 2 turkers agree on another response.

Thus, to reach an agreement with high accuracy, a minimum of 5 turkers is required with a QAT of 60% (3/5). The maximum number of repetitions should also be set as 5, because after starting with a QAT of 60%, the probability of getting an interturker agreement of 80% will then decrease and it will unnecessarily delay the item consensus calculation process if continuing to add more turkers to complete the task. In this study, therefore, 5 was set as both the minimum and maximum number of repetitions that each receipt received with a QAT of 80% (4/5), meaning that a receipt would continue to be available for turkers to interpret until it was assessed 5 times by 5 turkers and at least 4/5 (80%) turkers agreed on a response.

Before the formal launch of the MTurk tasks, several trial runs were performed using small subsets of pictures (ie, 10-20 receipt images at a time) to confirm that the tasks would be manageable by turkers and to confirm that the questions were easily understood and completed. While doing assignments, turkers could reach out to the requester directly by email or through the *Report a Problem with this Task* tab on the MTurk survey interface. During the trial tasks, we received several email inquiries from turkers regarding uncertainties about numbers versus letters in receipt identifiers, receipt number cutoff from the image at the top, etc. No turker expressed problems with interpreting other receipt information, either during the trial runs or the formal task.

Figure 2. Screenshot of the Amazon Mechanical Turk consensus task.



1. **What restaurant is this receipt from?**

○ McDonald's
○ Wendy's
○ Subway
○ Burger King
○ KFC
○ Other

2. **Is there a drink included on the receipt?**

○ Yes
○ No

3. **What type of beverage is included?**

○ cold beverage
○ hot beverage

4. **What is the receipt identifier? It is usually at the top and should be in the format of L####-###, where L is a letter and the #s are almost always NUMBERS, and there's a hyphen (not underscore) between them. Note, always capitalize the first letter, even if it's not capitalized when written.**

Guess if you can, or just put in the first letter if you can't. (The first letter is from the name of the store). If it's not visible at all, put in A0 (the letter A, followed by a zero)

Step 3: Evaluating the Reliability and Validity of Amazon Mechanical Turk

After all assignments were completed, we started evaluating the reliability and validity of MTurk for processing food purchase receipt information. This was conducted in three steps. First, interturker agreement was examined on responses to the four questions asked (ie, receipt identifier, restaurant name, beverage included or not, and type of beverage), and a majority response for each individual question was identified. Following that, turkers' majority responses were compared with gold standard values in the manually entered dataset to evaluate the reliability and validity of MTurk. Finally, we conducted sensitivity testing to assess whether and how the number of turkers completing each assignment would influence the agreement between turkers' majority responses and the gold standard. All analyses were conducted using R version 3.2.4 (R Foundation) [24].

Results

In total, 209 turkers participated in the *consensus* task and initiated or attempted 1346 assignments, among which 983 (73.03%) were approved or completed. On average, each turker contributed 4.7 assignments (SD 1.5). It took an average of 93.12 seconds (SD 70.5, median 65.0) for a turker to analyze a receipt; the entire project was completed within 40 minutes after we launched it on MTurk, with a total cost of US $80.80.

Table 1 lists the descriptive characteristics of the 196 prototypical receipts completed by 5 turkers. Among the 196 receipts that we sampled, one beverage item was listed in 140 receipts (71.4%), including 101 cold beverages (51.5%) and 39 hot beverages (19.9%). Among the 101 receipts with cold beverages, soda drinks were listed in 75 receipts (74.3%), including Coca-Cola, Sprite, Pepsi, Diet Coke, and generic drinks. The rest of the cold-beverage receipts included sweet tea (6/26, 23%), bottled water (5/26, 19%), milkshake or smoothie (5/26, 19%), iced coffee or coffee drinks (4/26, 15%), lemonade (3/26, 12%), and juice or juice beverages (3/26, 12%). Among the 39 receipts with hot beverages, hot coffee was listed in 36 receipts (92%), and the other 3 receipts included 1 with hot chocolate (3%) and 2 with hot tea (5%).

Table 1. Descriptive characteristics of the prototypical sample of receipts (N=196).

Restaurant	Number of receipts, n (%)	Number of receipts with beverages out of all receipts from the restaurant, n (%)	Number of hot beverages[a] out of all receipts from the restaurant, n (%)	Number of cold beverages[b] out of all receipts from the restaurant, n (%)
Burger King	15 (7.7)	12 (80)	0 (0)	12 (80)
KFC	40 (20.4)	29 (73)	8 (20)	21 (53)
McDonald’s	102 (52.0)	78 (76.5)	31 (30.4)	47 (46.1)
Subway	21 (10.7)	5 (24)	0 (0)	5 (24)
Wendy’s	18 (9.2)	16 (89)	0 (0)	16 (89)

[a] Hot beverages included hot coffee, hot chocolate, and hot tea.

[b] Cold beverages were mostly soda drinks, sweet tea, bottled water, and coffee drinks.

Turkers showed high agreement on their responses to the four questions that we asked. Specifically, among the 196 receipts that we sampled, the proportion of receipts with a QAT of at least 80% (ie, 4/5 interturker agreement) was 100% (196/196) for receipt identifier, 100% (196/196) for restaurant names (eg, Burger King, McDonald’s, or Subway), 98.5% (193/196) for beverage inclusion (ie, yes or no), 92.3% (181/196) for hot beverage (eg, hot coffee or hot tea), and 87.2% (171/196) for cold beverage (eg, soda or bottled water). At a QAT of 100%, the proportions of receipts with unanimous (ie, 5/5) agreement among the turkers was 100% (196/196) for receipt identifier, 90.8% (178/196) for restaurant names, 75.5% (148/196) for beverage inclusion, 69.4% (136/196) for hot beverages, and 51.0% (100/196) for cold beverages.

We further checked the disagreement pattern among turkers for specific questions. For the two questions on receipt identifiers and restaurant names, no disagreement was observed among turkers. When asked to indicate whether a beverage was included or not, disagreements started to emerge. For some cases, turkers overlooked beverages, especially soda drinks that were included in a combo rather than being listed as separate items. For others, some turkers wrongly categorized receipts with smoothies as *beverage not included*. Consequently, when it came to coding the specific type of beverage (ie, cold or hot beverage), more discrepancies were noted.

When comparing turkers’ majority responses with the gold standard data, the agreement rate was 100% (196/196) for receipt identifier, 100% (196/196) for restaurant name, 99.5% (195/196) for beverage inclusion, and 99.5% (195/196) for beverage types. We further tested whether and how the number of turkers influenced the agreement level between turkers’ majority responses and the gold standard data. Based on the analysis, when 3 turkers completed the project, the agreement between their consensus response and the gold standard data was 100% (196/196) for receipt identifier, 100% (196/196) for restaurant name, 99.5% (195/196) for beverage inclusion, and 99.5% (195/196) for beverage type, which were the same as the proportions when 5 turkers completed the assignments.

Discussion

This study is the first effort to assess whether MTurk, a popular crowdsourcing platform, can be used for processing data from food purchase receipts. In general, findings from this study supported the feasibility, reliability, and validity of MTurk as a cost-effective and time-efficient tool for processing food purchase receipt data.

Findings from this study demonstrated that, with minimal training, the MTurk workforce can categorize and analyze receipt data in a timely and cost-effective way. Despite the low compensation rate (ie, US $0.06 for every assignment), turkers in this study completed the entire task in less than 40 minutes, and the data extracted were of excellent quality, which was consistent with evidence from previous evaluation studies [9,13,18,25,26]. In fact, turkers in previous studies have expressed other motivations that enticed them to complete tasks. For example, many turkers felt it was a productive way to spend available free time, was mentally engaging, was oftentimes interesting, and offered a source of entertainment [27-30]. Compared with manual data extraction, which is often time-consuming, expensive, and difficult to scale up, MTurk can greatly enhance the widespread use of receipts as an assessment of food purchase and dietary behaviors.

Our findings further supported the reliability and validity of using MTurk for annotating receipt data, with high interrater agreement for both textual and multiple-choice questions. Previous studies have noted that as data coding tasks became more subjectively difficult, it got harder to achieve interpretive convergence [26]. Consistently in this study, we found perfect agreement (ie, 100%) among 5 turkers for the two easier questions that did not require subjective judgement (ie, receipt identifier and restaurant name), but we found increased disagreements for the two questions regarding beverage inclusion and beverage type. Nevertheless, when turkers’ majority responses were compared with the manually extracted gold standard, perfect or close-to-perfect agreements were observed, which confirmed the reliability of the number of 5 turkers that we requested for annotating individual receipts.

Our study has limitations. First, although we purposely selected receipts with clearly identifiable information, receipts used in health data analysis could sometimes be more difficult to read due to rips, pen markings, or small font, which would likely affect the agreement rates of turkers. Second, we only allowed turkers with prior approval ratings of 99% to participate in the task. Although this helped to ensure that turkers provide quality work, it also narrowed down the number of turkers available and likely increased the time for task completion. We did not test whether or how lowering the approval rating would affect the reliability and validity of data processing. Third, for demonstration purposes, the tasks we selected were objectively easy; future studies are warranted to determine if the same success rates can be obtained with more complicated tasks.

Despite the limitations, findings from this study hold important practical and research implications. First, our work confirmed the feasibility and accuracy of using MTurk as an innovative approach for processing data from food purchase receipts. In the future, the traditional model of manually annotating food purchase receipts as the gold standard for comparison may be flipped. Instead, crowdsourcing platforms could be used with appropriate task qualification requirements (eg, requiring turkers with prior approval ratings of 95% or 99%) to identify majority or consensus responses, followed by manually annotating a proportion of the receipts to confirm the reliability and validity. This feasibility study demonstrates the scalable and sustainable nature of this approach. Second, the accuracy of crowdsourced receipt annotation in this study lends strong support to the appropriateness of the number of turkers that we requested for each task. To get reliable consensus or majority responses among turkers when annotating image data on MTurk, we recommend future researchers set 5 as both the minimum and maximum number of repetitions for each image, with a question agreement threshold of 80%. Lastly, and most importantly, findings from this study point to the great potential of crowdsourcing for processing data in public health research, particularly tasks that cannot be entirely automated by computer programs and require human intelligence. A recent study has confirmed that objectively documented household food purchases from receipts can yield an unbiased and reasonably accurate estimate of overall diet quality as measured through 24-hour diet recalls [31]. With its time-saving and cost-effective qualities, crowdsourcing will vastly increase capacity for large-scale and high-quality receipt annotation, which in turn will advance our understanding of environmental influence on human health behaviors and ultimately lead to better health prevention and intervention efforts.

Acknowledgments

This study was funded by the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (grant number R01099241), the New York State Health Foundation (grant number 12-01682), and the Robert Wood Johnson Foundation (grant number 70823). The collaboration was supported by the American Academy of Health Behavior mentorship program.

Conflicts of Interest

None declared.

References

1. French SA, Shimotsu ST, Wall M, Gerlach AF. Capturing the spectrum of household food and beverage purchasing behavior: A review. J Am Diet Assoc 2008 Dec;108(12):2051-2058. [doi: 10.1016/j.jada.2008.09.001] [Medline: 19027408]
2. Drewnowski A, Rehm CD. Energy intakes of US children and adults by food purchase location and by specific food source. Nutr J 2013 May 08;12:59 [FREE Full text] [doi: 10.1186/1475-2891-12-59] [Medline: 23656639]
3. National Restaurant Association. 2016. National statistics: Restaurant industry facts at a glance URL: https://www.restaurant.org/News-Research/Research/Facts-at-a-Glance [accessed 2018-08-06] [WebCite Cache ID 71TLo9Nny]
4. French SA, Wall M, Mitchell NR, Shimotsu ST, Welsh E. Annotated receipts capture household food purchases from a broad range of sources. Int J Behav Nutr Phys Act 2009 Jul 01;6:37 [FREE Full text] [doi: 10.1186/1479-5868-6-37] [Medline: 19570234]
5. Gordon-Larsen P. Food availability/convenience and obesity. Adv Nutr 2014 Nov;5(6):809-817 [FREE Full text] [doi: 10.3945/an.114.007070] [Medline: 25398746]
6. Laska MN, Hearst MO, Forsyth A, Pasch KE, Lytle L. Neighbourhood food environments: Are they associated with adolescent dietary intake, food purchases and weight status? Public Health Nutr 2010 Nov;13(11):1757-1763 [FREE Full text] [doi: 10.1017/S1368980010001564] [Medline: 20529405]
7. White M. Food access and obesity. Obes Rev 2007 Mar;8 Suppl 1:99-107. [doi: 10.1111/j.1467-789X.2007.00327.x] [Medline: 17316311]
8. Elbel B, Kersh R, Brescoll VL, Dixon LB. Calorie labeling and food choices: A first look at the effects on low-income people in New York City. Health Aff (Millwood) 2009;28(6):w1110-w1121. [doi: 10.1377/hlthaff.28.6.w1110] [Medline: 19808705]
9. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing: Harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203 [FREE Full text] [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]
10. Mason W, Watts DJ. Financial incentives and the "performance of crowds". SIGKDD Explor 2009 Dec;11(2):100-108. [doi: 10.1145/1809400.1809422]
11. Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2018-06-18] [WebCite Cache ID 70GgopezZ]

12. Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: Challenges and opportunities. Brief Bioinform 2016 Jan;17(1):23-32 [FREE Full text] [doi: 10.1093/bib/bbv021] [Medline: 25888696]
13. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of crowdsourcing in health: Systematic review. J Med Internet Res 2018 May 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
14. Bohannon J. Psychology: Mechanical Turk upends social sciences. Science 2016 Jun 10;352(6291):1263-1264. [doi: 10.1126/science.352.6291.1263] [Medline: 27284175]
15. Saunders DR, Bex PJ, Woods RL. Crowdsourcing a normative natural language dataset: A comparison of Amazon Mechanical Turk and in-lab data collection. J Med Internet Res 2013 May 20;15(5):e100 [FREE Full text] [doi: 10.2196/jmir.2620] [Medline: 23689038]
16. Yu B, Willis M, Sun P, Wang J. Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk. J Med Internet Res 2013 Jun 03;15(6):e108 [FREE Full text] [doi: 10.2196/jmir.2513] [Medline: 23732572]
17. Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. J Med Internet Res 2013 Jul 31;15(7):e144 [FREE Full text] [doi: 10.2196/jmir.2569] [Medline: 23903235]
18. Kuang J, Argo L, Stoddard G, Bray BE, Zeng-Treitler Q. Assessing pictograph recognition: A comparison of crowdsourcing and traditional survey approaches. J Med Internet Res 2015 Dec 17;17(12):e281 [FREE Full text] [doi: 10.2196/jmir.4582] [Medline: 26678085]
19. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med 2014 Feb;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]
20. Ilakkuvan V, Tacelosky M, Ivey KC, Pearson JL, Cantrell J, Vallone DM, et al. Cameras for public health surveillance: A methods protocol for crowdsourced annotation of point-of-sale photographs. JMIR Res Protoc 2014 Apr 09;3(2):e22 [FREE Full text] [doi: 10.2196/resprot.3277] [Medline: 24717168]
21. Cantor J, Torres A, Abrams C, Elbel B. Five years later: Awareness of New York City's calorie labels declined, with no changes in calories purchased. Health Aff (Millwood) 2015 Nov;34(11):1893-1900. [doi: 10.1377/hlthaff.2015.0623] [Medline: 26526247]
22. Dowse R, Ehlers MS. The evaluation of pharmaceutical pictograms in a low-literate South African population. Patient Educ Couns 2001 Nov;45(2):87-99. [Medline: 11687321]
23. Kim H, Nakamura C, Zeng-Treitler Q. Assessment of pictographs developed through a participatory design process using an online survey tool. J Med Internet Res 2009 Feb 24;11(1):e5 [FREE Full text] [doi: 10.2196/jmir.1129] [Medline: 19275981]
24. R Project. The R Project for statistical computing URL: https://www.r-project.org/ [WebCite Cache ID 71TPlKnhO]
25. Azzam T, Jacobson MR. Finding a comparison group. Am J Eval 2013 Jun 18;34(3):372-384. [doi: 10.1177/1098214013490223]
26. Mitra T, Hutto C, Gilbert E. Comparing person- and process-centric strategies for obtaining quality data on Amazon Mechanical Turk. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea p. 1345-1354. [doi: 10.1145/2702123.2702553]
27. Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and fast: But is it good? Evaluating non-expert annotations for natural language tasks. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: 2008 Conference on Empirical Methods in Natural Language Processing; October 25-27, 2008; Honolulu, Hawaii p. 254-263 URL: http://www.aclweb.org/anthology/D08-1027
28. Chandler D, Kapelner A. Breaking monotony with meaning: Motivation in crowdsourcing markets. J Econ Behav Organ 2013 Jun;90:123-133. [doi: 10.1016/j.jebo.2013.03.003]
29. Horton JJ, Rand DG, Zeckhauser RJ. The online laboratory: Conducting experiments in a real labor market. Exp Econ 2011 Sep;14(3):399-425. [doi: 10.1007/s10683-011-9273-9]
30. Paolacci G, Chandler J, Ipeirotis P. Running experiments on Amazon Mechanical Turk. Judgm Decis Mak 2010 Aug;5(5):411-419 [FREE Full text]
31. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? Int J Behav Nutr Phys Act 2017 Dec 11;14(1):46 [FREE Full text] [doi: 10.1186/s12966-017-0502-2] [Medline: 28399887]

Abbreviations

HIT: Human Intelligence Task
MTurk: Amazon Mechanical Turk
QAT: question agreement threshold

Edited by G Eysenbach; submitted 27.08.18; peer-reviewed by P Liu, FA Shafie; comments to author 06.10.18; revised version received 15.12.18; accepted 16.12.18; published 05.04.19.

Please cite as:
Lu W, Guttentag A, Elbel B, Kiszko K, Abrams C, Kirchner TR
Crowdsourcing for Food Purchase Receipt Annotation via Amazon Mechanical Turk: A Feasibility Study
J Med Internet Res 2019;21(4):e12047
URL: http://www.jmir.org/2019/4/e12047/
doi:10.2196/12047
PMID:30950801

Original Paper

Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study

Nassr Nama[1,2,3,4], MD, MSc; Margaret Sampson[2], MLiS, PhD, AHIP; Nicholas Barrowman[1,2,5], PhD; Ryan Sandarage[3], BSc; Kusum Menon[1,2], MD, MSc; Gail Macartney[2], RN(EC), PhD; Kimmo Murto[1,6], MD; Jean-Philippe Vaccani[1,7], MD; Sherri Katz[1,2], MDCM, MSc; Roger Zemek[1,2,8], MD; Ahmed Nasr[1,9], MD, PhD; James Dayre McNally[1,2], MD, PhD

[1]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[2]Department of Pediatrics, Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada
[3]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[4]Department of Pediatrics, British Columbia Children's Hospital, Vancouver, BC, Canada
[5]Clinical Research Unit, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
[6]Department of Anesthesiology and Pain Medicine, Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada
[7]Department of Otolaryngology, Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada
[8]Department of Emergency Medicine, Faculty of Medicine, Ottawa, ON, Canada
[9]Division of Pediatric Surgery, Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada

Corresponding Author:
James Dayre McNally, MD, PhD
Department of Pediatrics
Children’s Hospital of Eastern Ontario
401 Smyth Road
Ottawa, ON,
Canada
Phone: 1 6137377600 ext 3553
Email: dmcnally@cheo.on.ca

Abstract

Background: Systematic reviews (SRs) are often cited as the highest level of evidence available as they involve the identification and synthesis of published studies on a topic. Unfortunately, it is increasingly challenging for small teams to complete SR procedures in a reasonable time period, given the exponential rise in the volume of primary literature. Crowdsourcing has been postulated as a potential solution.

Objective: The feasibility objective of this study was to determine whether a crowd would be willing to perform and complete abstract and full text screening. The validation objective was to assess the quality of the crowd’s work, including retention of eligible citations (sensitivity) and work performed for the investigative team, defined as the percentage of citations excluded by the crowd.

Methods: We performed a prospective study evaluating crowdsourcing essential components of an SR, including abstract screening, document retrieval, and full text assessment. Using CrowdScreenSR citation screening software, 2323 articles from 6 SRs were available to an online crowd. Citations excluded by less than or equal to 75% of the crowd were moved forward for full text assessment. For the validation component, performance of the crowd was compared with citation review through the accepted, gold standard, trained expert approach.

Results: Of 312 potential crowd members, 117 (37.5%) commenced abstract screening and 71 (22.8%) completed the minimum requirement of 50 citation assessments. The majority of participants were undergraduate or medical students (192/312, 61.5%). The crowd screened 16,988 abstracts (median: 8 per citation; interquartile range [IQR] 7-8), and all citations achieved the minimum of 4 assessments after a median of 42 days (IQR 26-67). Crowd members retrieved 83.5% (774/927) of the articles that progressed to the full text phase. A total of 7604 full text assessments were completed (median: 7 per citation; IQR 3-11). Citations from all but 1 review achieved the minimum of 4 assessments after a median of 36 days (IQR 24-70), with 1 review remaining incomplete after 3 months. When complete crowd member agreement at both levels was required for exclusion, sensitivity was 100% (95% CI 97.9-100) and work performed was calculated at 68.3% (95% CI 66.4-70.1). Using the predefined alternative 75% exclusion threshold, sensitivity remained 100% and work performed increased to 72.9% (95% CI 71.0-74.6; *P*<.001). Finally, when a simple majority threshold was considered, sensitivity decreased marginally to 98.9% (95% CI 96.0-99.7; *P*=.25) and work performed increased substantially to 80.4% (95% CI 78.7-82.0; *P*<.001).

Conclusions: Crowdsourcing of citation screening for SRs is feasible and has reasonable sensitivity and specificity. By expediting the screening process, crowdsourcing could permit the investigative team to focus on more complex SR tasks. Future directions should focus on developing a user-friendly online platform that allows research teams to crowdsource their reviews.

(J Med Internet Res 2019;21(4):e12953) doi:10.2196/12953

KEYWORDS

crowdsourcing; systematic reviews as topic; meta-analysis as topic; research design

Introduction

Systematic Reviews and their Challenges

Systematic reviews (SRs) are often cited as the highest level of evidence available as they involve the identification and synthesis of all published studies on a topic [1]. Moreover, given the rise in the volume of primary literature, clinicians, scientists, and policy makers increasingly rely on SRs to inform decision making on important issues [2]. Maintenance of a continuous stream of up-to-date, high-quality evidence is important for optimal patient care and proper utilization of health care resources [3-7]. Unfortunately, it is more and more challenging for individuals and small teams to complete SR procedures in a reasonable time period [8-11]. To complete an SR, investigators need to manage thousands of potentially relevant citations, remove duplicates, screen abstracts for eligibility, download manuscripts, independently review full texts, resolve conflicts regarding eligibility, assess quality, extract and analyze data, and author a manuscript [8]. Consequently, there is significant interest in novel methodological approaches that improve the feasibility and completion of knowledge synthesis efforts and also avoid the scenario where investigators choose less than optimal search and screening strategies to maintain feasibility [12,13].

Crowdsourcing in Science

Crowdsourcing has been postulated as a potential solution to address the barriers to efficient completion of SRs [14]. Crowdsourcing is “the practice of obtaining participants, services, ideas, or content by soliciting contributions from a large group of people, especially via the Internet” [15,16]. From tracking soil quality [17] and classifying galaxies [18] to identifying the three-dimensional (3D) configuration of complex protein structures [19], crowdsourcing has been studied and validated in other scientific areas. More recently, the medical field has seen increased application of crowdsourcing approaches to a wide range of problems ranging from funding research [20] to disease diagnosis (eg, Cell Slider [21]). In recent years, a small number of research groups have proposed and even evaluated crowdsourcing certain SR tasks [14,22-24]. These studies mainly focused on abstract screening, and to our knowledge, no previous research has studied the crowd’s capacity for full text retrieval and review.

Objectives

The feasibility objective of this study was to determine whether it was possible to recruit an online crowd to perform and complete abstract and full text screening for SRs. The validation objective was to assess the quality of work performed by the crowd when compared with the gold standard expert approach, both in regards to the sensitivity for eligible citations and the potential work performed for the investigative team.

Methods

Study Design

This study was conducted at the Children’s Hospital of Eastern Ontario (CHEO), a teaching hospital affiliated with the University of Ottawa. Similar to previous studies in this field [22], and as per the CHEO Research Ethics Board, this study was not considered as research on humans, and as such, ethics approval was not required. The project description clearly stated that crowd members were not eligible for authorship and that their contribution was part of a research study validating crowdsourcing as a new methodology in the area of SRs. On sign-up and log-in, the crowd was provided with both privacy policy and terms of use documentation, designed in consultation with the CHEO privacy lawyer (Multimedia Appendix 1).

This study was a prospective quantitative study evaluating the feasibility and validity of crowdsourcing essential components of an SR, including abstract screening, document retrieval, and full text assessment. For the validation component, performance of the online volunteer crowd was compared with citation review through the accepted, gold standard, trained expert approach. Results are reported according to the Standards for Reporting of Diagnostic Accuracy Studies guidelines for diagnostic accuracy studies [25] (Multimedia Appendix 2).

Study Outcomes

The primary outcome for the feasibility component was the number of citations that achieved the target number of independent assessments. Consistent with our initial pilot study, feasibility success was a priori defined as achieving a minimum of 4 independent assessments per citation [23]. The primary outcome for the validation component was the ability of the crowd to identify and retain eligible studies at the abstract level (sensitivity). For the validation component, secondary outcomes included the crowd’s overall sensitivity after full text review of retained abstracts and the work performed. Work performed was defined as the percentage of all citations that were excluded by the crowd and did not require assessment by the investigative team at abstract or full text levels. To allow comparison with other studies, specificity was also calculated. Individual reviewer's performance represented an exploratory outcome.

Sample Size and Power

For the purpose of the sample size calculation, the crowd retention of true positives (ie, the sensitivity) was assumed to be 95% at the abstract screening level. Under this assumption, the sample size was selected so that the lower end of the 95% CI for sensitivity would be no less than 90%. Using a Wilson score CI, this would be the case if 142 abstracts were retained by crowd members out of a total of 150 abstracts deemed eligible by expert screeners (95% sensitivity). Thus, a sample size of 150 abstracts was selected.

Systematic Review Selection and Details

Potentially eligible SRs included those initiated during 2016 and not anticipated to be published before the end of the 2017 calendar year (to prevent crowd members from accessing the published data with lists of eligible papers). The reviews selected covered the areas of anesthesiology, cardiology, emergency medicine, endocrinology [13], respirology, and general surgery [26] (Table 1). We targeted a wide range of topics with the intention of making the results more generalizable and increasing the likelihood that a potential crowd participant would identify a topic of interest. For each SR, the principal investigator was asked to provide the following: (1) inclusion and exclusion criteria and (2) the final list of citations determined to be eligible by their expert reviewers (true positives). In some circumstances, the investigative team provided screening criteria that differed slightly from their original review. In this circumstance, study authors NN and DM reviewed the true positives against criteria presented to the crowd, and any study not meeting the eligibility criteria provided to the crowd was removed from the true positive list. For SRs exceeding 1000 citations, smaller subsets were chosen, ensuring a reasonable pool of true positives (Table 1).

Crowd Recruitment and Compensation

To qualify for participation, the individual needed to be both a nonexpert and a member of a large distributed crowd. To be considered a nonexpert, individuals had to confirm that they had not participated in the development of the protocol for the SR and had not received training sessions by the investigators on how to screen citations. For this initial feasibility study, we targeted the large online crowd or population of individuals with some postsecondary or postgraduate training, including undergraduate, medical students, residents, nurses, and other allied health specialists. We targeted this population for 2 reasons: (1) given the paucity of work on crowdsourcing SRs, it seemed appropriate to begin by evaluating the performance measures in a cohort with or receiving applicable science or health training and (2) similarly skilled and motivated individuals would be available and accessible at dozens of cities in Canada and hundreds through the world. Individuals were recruited by sending emails to (1) the hospital volunteer department, (2) University of Ottawa Medical School, (3) student interest groups at the 17 Canadian medical schools, and (4) health-related undergraduate student groups in 22 universities across Canada. Promotional material was designed by CHEO Media House (Multimedia Appendix 3). As a resource for those who might want to recruit a crowd with similar characteristics and motives to perform a large SR, we have provided an example copy of the email sent (Multimedia Appendix 3). Compensation was limited to the potential for a gift card (Can $100) for the top 3 crowd members in each review (highest number of citations screened accurately). Furthermore, we offered crowd members the possibility to connect them with CHEO investigators performing an SR and seeking to grow their research team. For reviews that did not attain the minimum of 4 assessments per citation at either level after 2 months, additional gift cards were offered. In total, 26 gift cards were distributed among 22 crowd members.

Crowdsourcing Website Development and Overview of the Platform Function

To complete this study, we used the CrowdScreenSR citation screening platform, as previously described [23]. The website was adopted by the CHEO Research Institute in 2016 and was concurrently used by 4 of the 6 investigative teams for completion of their SRs using the gold standard or expert approach. Crowd members had unique usernames and passwords, allowing separate tracking and evaluation of progress, work performed, and performance. Demographic data were collected on crowd members, including the level of training, research experience, participation in previous SRs, and number of publications. Crowd members were instructed to select only the highest level of training in progress or completed. Initially, each crowd member was given access to a demonstration module to help familiarize them with the website functioning. Initially, all 6 reviews were shown to the crowd, along with a description of the goal of the study and its eligibility criteria. For each SR, a training set of 10 citations, including 2 to 3 true positives, was used to familiarize the crowd with both the SR eligibility criteria and platform. During this training set, immediate feedback was provided on whether the crowd member's assessment of the citation was accurate. Crowd members who completed the training set were given access to the full set of citations for that review (regardless of their performance). A minimum goal of at least 50 citations was set, with crowd members offered the flexibility of screening as many citations as desired. For both abstract and full text screening levels, the crowd members were instructed to place citations into 1 of the 3 groups: (1) retain, (2) exclude, or (3) no assessment (not comfortable assessing this citation). When a citation was categorized as exclude, the crowd member was further prompted to provide which eligibility criteria were not met. We have aimed to achieve at least four assessments per citation at each of abstract and full text levels, with no predefined maximum. Abstract-level screening started on January 7, 2017, and was completed on April 23, 2017. Retrieval of manuscripts, PDF upload, and full text screening continued until September 3, 2017. Start dates for each of these phases were chosen at the beginning of university trimesters to maximize crowd members participation.

Table 1. Description of systematic reviews.

Systematic review[a]	Description	Total citations[b], N	Validation study[c], N	Eligible citations[d], N (%)
Anesthesiology[e]	A systematic review of preoperative screening for factors associated with postoperative critical respiratory events in children undergoing elective adenotonsillectomy	5458	300	29 (9.7)
Cardiology[f]	A scoping review of all randomized controlled trials in pediatric cardiology	7540	490	71 (14.5)
Emergency	A systematic review of studies on concussion education and outcomes for children	513	503	9 (1.8)
Endocrinology[g]	2017 update of a previously published systematic review on high-dose supplementation of vitamin D in children [13,23]	201	201	30 (14.9)
Respirology	A systematic review of studies on predictors of positive airway pressure adherence at home among children with sleep-disordered breathing	277	265	23 (8.7)
Surgery	A systematic review of studies on asymptomatic antenatal diagnoses of congenital pulmonary airways malformation that describe natural history of the disease and future symptoms [26]	574	564	16 (2.8)

[a] Total of 6 systematic reviews and 2323 citations were included. 178 (7.7%) of citations were identified as eligible by the experts (ie, true positives).

[b] Total number of citations identified by the search strategy.

[c] Number of citations included in the validation study, after excluding the 10 citations used as a training set.

[d] Eligible citations as identified by the experts (ie, true positives).

[e] A random sample of 300 citations was selected and enriched with up to 30 eligible citations.

[f] A random sample of 500 citations was selected.

[g] Given the limited number of citations, the 10 training set citations were selected from the original publication.

Advancement to Full Text Screening

To focus on the crowd's capacity to assess abstracts, citations with missing abstracts were automatically pushed forward to full text retrieval and review. In addition to those with missing abstracts, citations where greater than or equal to 25% of the crowd assessed as eligible were retained for use in the assessment of crowd performance at full text review (Figure 1).

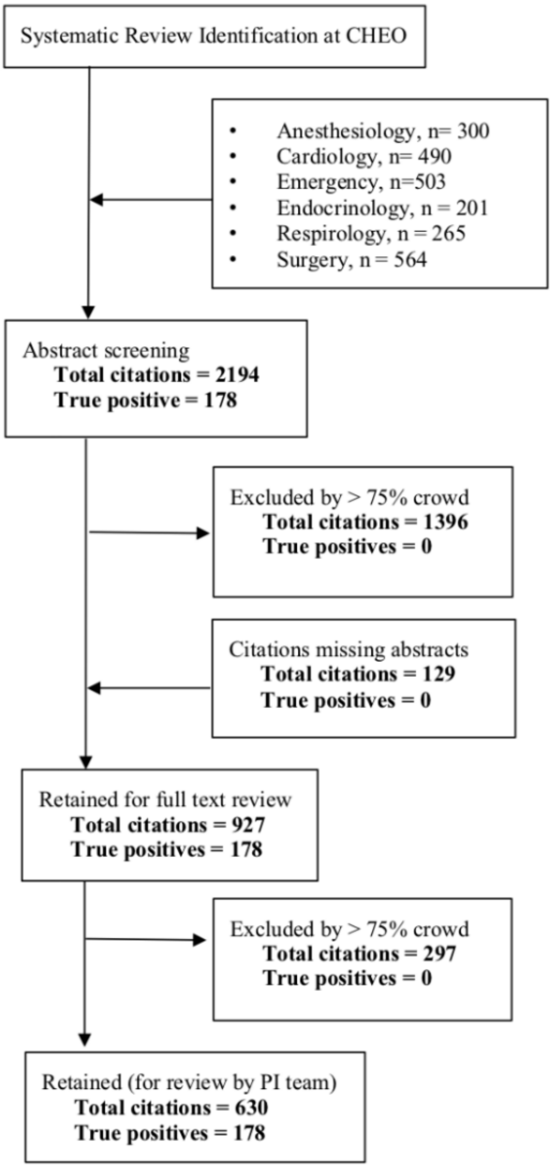
Validation of Crowd Performance

For the purpose of the analysis, different thresholds for citations' exclusion were tested. Specific exclusion cut-offs (75% and 100%) were prioritized in the analysis as these were both tested and performed well in our previous study [23]. To allow comparison with another recently published study [22], we also considered the 50% exclusion threshold (Multimedia Appendix 4). Finally, a range of exclusion thresholds between 0% and 100% were tested and presented graphically. Using the 0% cut-off was the least conservative approach, where a citation was excluded if any crowd member opted to exclude. On the other end, the 100% cut-off was the most conservative, and a citation was only excluded if every crowd member chose to exclude. Measures of individual crowd members' performance were completed as an exploratory analysis and were limited to those crowd members having completed a minimum of 50 citation assessments. This cut-off was established a priori, as crowd members were asked to complete a minimum of 50 citations to increase the chance that the subset of citations assessed would contain at least a few eligible papers.

Data Analysis

Data analysis was performed using SAS (version 9.4; SAS Institute, Cary, NC, USA). Figures were generated using GraphPad Prism (version 8.0; GraphPad Software, Inc, La Jolla, CA, USA). Fisher exact and Pearson Chi-square tests were used to compare characteristics of crowd members who proceeded to complete the minimum 50 citations with those who did not. Wilson score method was used to calculate 95% CIs for sensitivity, specificity, and work performed. The McNemar 1-tailed test was used to compare sensitivity and work performed between different exclusion thresholds. As a more stringent threshold for excluding a paper can only increase the sensitivity, a 1-tailed test was used.

Figure 1. Study flow diagram. To focus the study on the crowd's capacity to assess abstracts and not title screening, citations with missing abstracts (129) were removed. These citations were later added to the full text screening stage, along with any citation that did not receive higher than our a priori exclusion threshold of 75% at the abstract screening level. True positives reflect the number of citations that were identified as eligible by the experts. CHEO: Children's Hospital of Eastern Ontario; PI: principal investigator.



Results

Crowd Description

A total of 313 individuals signed up on the CrowdScreenSR website. None of those were deemed ineligible based on our criteria. Of the 312 potential crowd members, 171 (54.8%) initiated at least one SR training set and 117 (37.5%) completed the training set and commenced abstract screening. Of these 117 crowd members, 71 (60.7%) completed 50 or more independent citations (Table 2). With regards to the crowd's demographics, the most commonly selected answers were the highest level of training as undergraduate studies (131/312, 42.0%) and some prior research experience (220/312, 70.5%). One-third of the participants reported having been an author on a least one research publication (103/312, 33.0%), with only 1 in 5 citing previous involvement with SR research (65/312, 20.8%). Comparing those who proceeded to complete the minimum 50 citations with those who did not showed no statistically significant differences with respect to the level of training, prior research experience, publications of any kind, or involvement in SRs.

Table 2. Comparison of crowd members who proceeded to complete the minimum 50 citations with those who did not.

Crowd members	<50 assessments[a], N (%)	≥50 assessments, N (%)	*P* value[b]	Total
Total reviewers	241 (77.2)	71 (22.8)	__[c]	312
Background[d]	—	—	.15	—
Undergraduate studies	107 (44.4)	24 (33.8)	—	131
Medical student	41 (17.0)	20 (28.2)	—	61
Graduate studies	36 (14.9)	9 (12.7)	—	45
Allied health professional	20 (8.3)	3 (4.2)	—	23
Physician	7 (2.9)	3 (4.2)	—	10
Other	4 (1.7)	3 (4.2)	—	7
Research involvement[e]	—	—	.08	—
None	65 (27.0)	27 (38.0)	—	92
Student	130 (53.9)	35 (49.3)	—	165
Volunteer	81 (33.6)	23 (32.4)	—	104
Coordinator	66 (27.4)	11 (15.5)	—	77
Investigator	25 (10.4)	3 (4.2)	—	28
Publications	—	—	.23	—
None	156 (64.7)	53 (74.6)	—	209
1-3	57 (23.7)	14 (19.7)	—	71
>3	28 (11.6)	4 (5.6)	—	32
Systematic reviews experience	—	—	—	—
Involvement in a review	52 (21.6)	13 (18.3)	.62	65
Leading a review	12 (5.0)	5 (7.0)	.55	17
Publishing a review	38 (15.8)	12 (16.9)	.85	50

[a] Minimum of 50 citations in a systematic review was requested from crowd members at the beginning of the study. Crowd members with 50 citations or more performed 98.8% (16,789/16,988) and 93.0% (7071/7604) of the abstract and full text assessments, respectively.

[b] Comparison between those who did less than 50 assessments and those who did 50 or more (Fisher test).

[c] Not applicable.

[d] Only 277 crowd members provided their background.

[e] Multiple choices can be selected by reviewers.

Systematic Review Tasks Performed by Crowd (Feasibility)

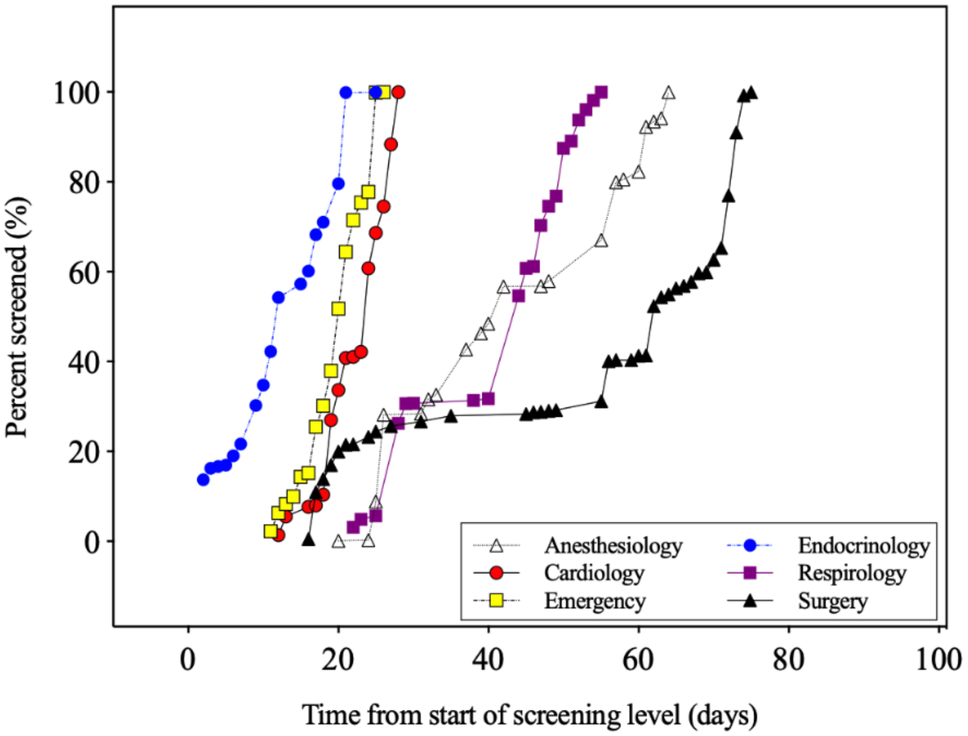
Abstract Screening

Crowd members performed 16,988 abstract assessments on 2194 unique citations, and all of the citations met or exceeded the feasibility target of 4 independent assessments, with a median of 8 assessments per paper (interquartile range [IQR] 7-8; Multimedia Appendix 5). The time required to acquire 4 independent assessments per citation at the abstract level varied by review, with a median of 42 days (IQR 26-67; Figure 2). A total of 3 reviews were completed in less than a month: endocrinology (25 days), emergency (26 days), and cardiology (28 days). Of the remaining reviews, 2 (respirology and anesthesiology) required 55 and 64 days for completion, respectively. A total of 2 months after the project was launched, the 1 remaining review remained below the target of 4 assessments per paper (<50%). When the incentive was revised to a Can $100 gift card for any crowd member that completed all citations in that review, the project was completed in the subsequent 14 days.

Retrieval of Full Text

Following abstract screening, 927 papers were pushed to the full text level (Figure 1). Crowd members were able to successfully retrieve 83.5% (774/927) of the articles that progressed to the full text review phase. Of the 153 articles that were not retrieved by the crowd, 95% (145) were not open access and not available through the University of Ottawa.

Figure 2. Time to review completion during abstract screening. Time required to complete the desired 4 assessments per citation at the abstract screening level. On day 61, additional incentives were offered for the surgery review.



Full Text Screening

At full text review, the crowd members performed 7604 assessments on 927 unique articles (Multimedia Appendix 5). Of the 6 SRs, 5 achieved 4 assessments for all of their citations. Overall, median assessments number per citation was 7 (IQR 3-11). Full text review required a median of 36 days (IQR 24-70), with 1 review that remained incomplete after 3 months (Figure 3). In the first month, the crowd completed both the cardiology (23 days) and endocrinology (24 days), with emergency completed shortly thereafter (36 days). Anesthesiology was at 89.3% at the time and remained without significant progression until an email notifying the crowd that the remaining reviews were closing was sent on day 60, and the review was completed a day later. Additional $100 gift cards were offered for the other 2 SRs. With these efforts, the respirology review was completed after 79 days. The crowd did not complete full text review for the surgery SR, with only 1.4% (4/283) of citations above the 4-assessments threshold; 82 citations of those had only 2 assessments, and the other 197 citations had 3 assessments at the full text stage.

Validation of the Crowd Performance—Abstract Level

When complete crowd member agreement at the abstract level was required for exclusion, sensitivity was 100% (95% CI 97.9-100) and work performed was calculated at 44.9% (95% CI 42.8-46.9; Table 3). Using the predefined 75% exclusion threshold, with citations excluded if more than 75% of the crowd agreed at the abstract level, sensitivity remained 100% and the work performed increased to 60.1% (95% CI: 58.1-62.1; *P*<.001). Finally, when a simple majority was required to exclude a citation, sensitivity decreased marginally to 98.9% (95% CI 96.0-99.7; *P*=.25) and the work performed increased to 68.0% (95% CI 66.1-69.9; *P*<.001). Sensitivity and work performed data were calculated for each of the individual SRs (Multimedia Appendix 6). Crowd specificity for abstract screening at 100%, 75%, and 50% exclusion thresholds was calculated as 48.6%, 65.1%, and 73.6%, respectively. Finally, the relationship between sensitivity and work performed after abstract screening at exclusion thresholds ranging from 0% to 100% is presented in Figure 4.

Validation of the Crowd Performance—Full Text Level

Crowd’s performance was assessed after full text screening of retained abstracts. All eligible citations that were retained at the abstract level were also retained by the crowd at the full text level, and sensitivity remained the same based on the 3 exclusion thresholds (Table 3). When complete crowd member agreement at both levels was required for exclusion, work performed was calculated at 68.3% (95% CI 66.4-70.1). Using the predefined 75% exclusion threshold, with citations excluded if more than 75% of the crowd agreed at both the abstract and full text level, the work performed increased to 72.9% (95% CI 71.0-74.6; *P*<.001; Table 3). Finally, when a simple majority was required to exclude a citation, the work performed increased substantially to 80.4% (95% CI 78.7-82.0; *P*<.001). Sensitivity and work performed after screening both levels were calculated for each of the individual SRs (Multimedia Appendix 7). Crowd specificity after screening both levels at the 100%, 75%, and 50% exclusion thresholds were calculated as 73.9%, 78.9%, and 87.0%, respectively. Finally, the relationship between sensitivity and work performed at exclusion thresholds ranging from 0% to 100% is presented in Figure 5.

Figure 3. Time to review completion during full text screening. Time required to complete the desired 4 assessments per citation at the full screening level. Between days 58 and 77, reviewers were notified that the screening deadline is for day 90, and further incentives were offered for the anesthesiology, surgery and respirology reviews.

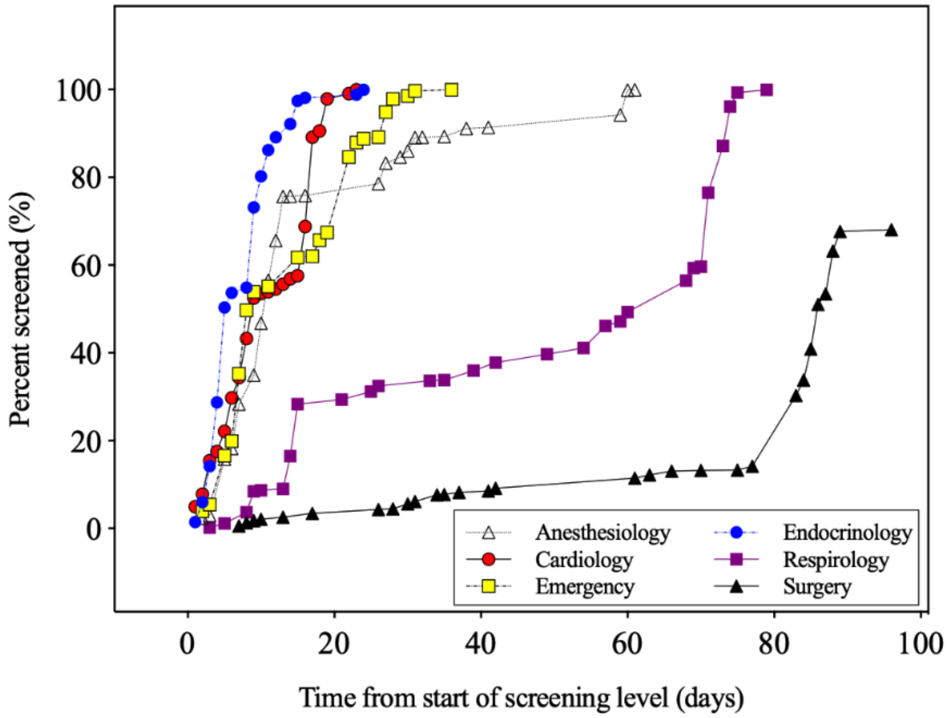


Table 3. Crowd's sensitivity and work performed at different exclusion thresholds.

Crowd agreement required to exclude[a]	Sensitivity[b]		Work performed[c]		Specificity[d]	
	Mean (95% CI)	*P* value[e]	Mean (95% CI)	*P* value[e]	Mean (95% CI)	*P* value[e]
Abstract level[f]						
=100%	100 (97.9-100)	.50	44.9 (42.8-46.9)	<.001	48.6 (46.5-50.7)	<.001
>75%	100 (97.9-100)	(Ref[g])	60.1 (58.1-62.1)	(Ref)	65.1 (63.0-67.1)	(Ref)
>50%	98.9 (96.0-99.7)	.25	68.0 (66.1-69.9)	<.001	73.6 (71.7-75.4)	<.001
Full text level[h]						
=100%	100 (97.9-100)	.50	68.3 (66.4-70.1)	<.001	73.9 (72.0-75.8)	<.001
>75%	100 (97.9-100)	(Ref)	72.9 (71.0-74.6)	(Ref)	78.9 (77.2-80.6)	(Ref)
>50%	98.9 (96.0-99.7)	.25	80.4 (78.7-82.0)	<.001	87.0 (85.5-88.4)	<.001

[a] Citations were excluded based on different thresholds.

[b] Sensitivity is the percentage of eligible citations, identified by the experts, that were retained by the crowd.

[c] Work performed is the percentage of citations that were excluded by the crowd and did not require assessment by the investigative team at the abstract level.

[d] Specificity is the percentage of ineligible citations, as identified by the experts, that were excluded by the crowd.

[e] *P* value compares sensitivity, work performed, or specificity to the respective value at the 75% threshold (McNemar test).

[f] Outcomes were measured after abstract screening. A citation was excluded if the percentage of assessments that excluded the paper at the abstract level was higher than the specified threshold.

[g] Ref: reference category.

[h] Outcomes were measured at the end of both screening levels. A citation was excluded if the percentage of assessments that excluded the paper at either abstract or full text levels was higher than the specified threshold.

Figure 4. Sensitivity and work performed as a function of the exclusion threshold at the abstract level. A citation is excluded when the percentage of exclusion assessment is above the exclusion cut-off at the abstract level. Sensitivity is the percentage of eligible citations identified by the experts that were retained by the crowd. Work performed is the percentage of citations that were excluded by the crowd and did not require assessment by the investigative team at the abstract level.

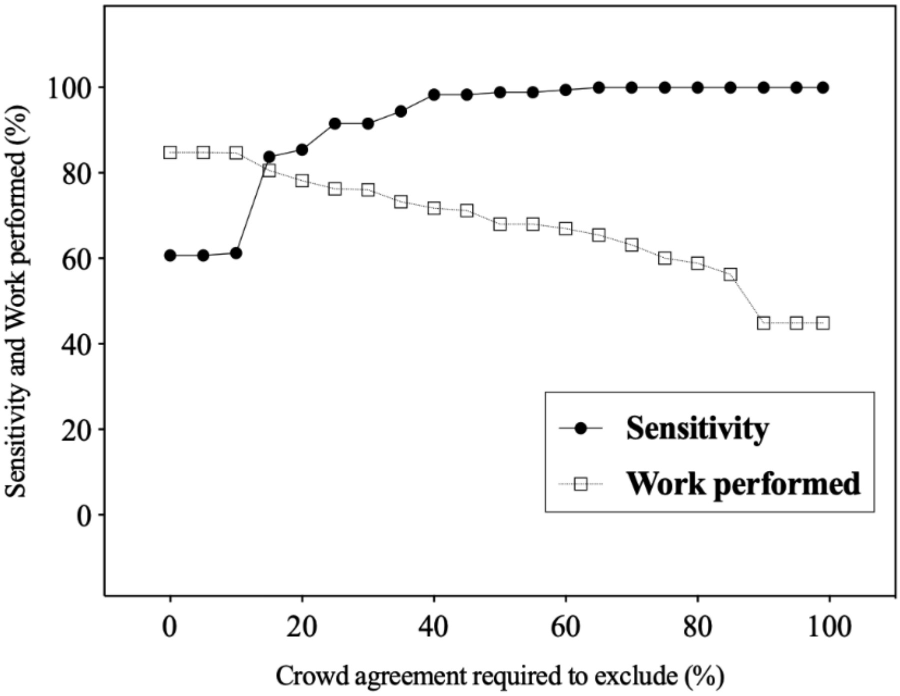


Figure 5. Sensitivity and work performed as a function of the exclusion threshold after abstract and full text screening. A citation is excluded when the percentage of exclusion assessment is above the exclusion cut-off at either abstract or full text screening. Sensitivity is the percentage of eligible citations identified by the experts that were retained by the crowd. Work performed is the percentage of citations that were excluded by the crowd and did not require assessment by the investigative team at abstract or full text levels.

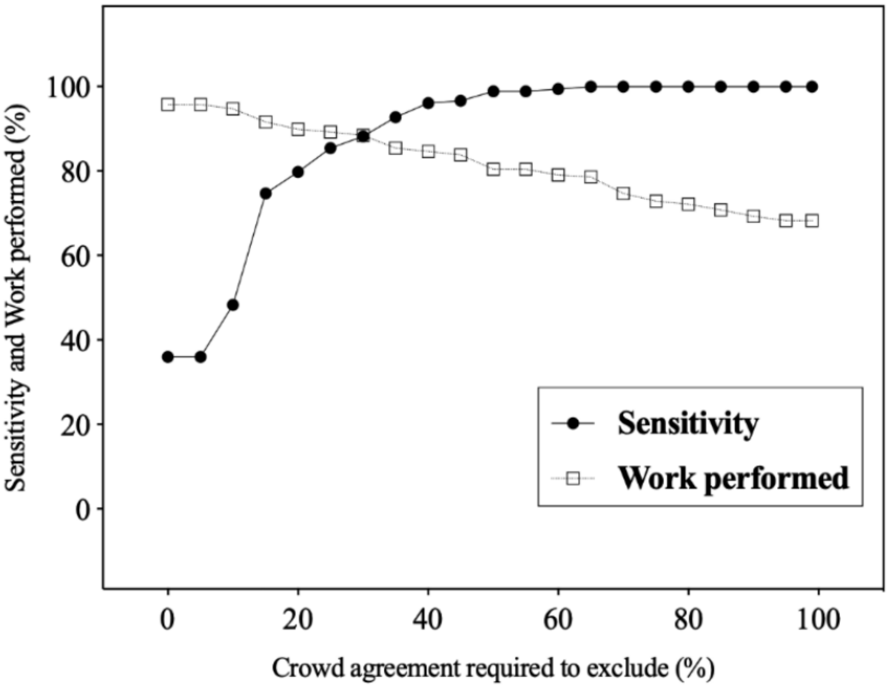


Table 4. Individual crowd members' performance.

Performance[a,b]	Abstract level (N=40)		Full text level (N=41)	
	Median (IQR[c])	Range	Median (IQR)	Range
Assessments	306.5 (108.5-513.5)	16-2194	141 (72-206)	5-786
Sensitivity[d]	96.6 (92.0-100.0)	55.0-100.0	96.7 (89.6-99.0)	32.3-100.0
Specificity[e]	76.4 (66.2-92.8)	42.4-96.3	64.3 (58.5-73.8)	22.9-100.0

[a] Only crowd members who have completed 50 assessments or more in 1 review were included in this table. Crowd members with 50 citations or more performed 98.8% (16,789/16,988) and 93.0% (7071/7604) of the abstract and full text assessments, respectively.

[b] Results are provided per crowd member.

[c] IQR: interquartile range.

[d] Sensitivity is the percentage of eligible citations, identified by the experts, that were retained by the crowd member. It is based on 38 crowd members at the abstract level and 38 at the full text level. The remaining crowd members did not assess any eligible citations.

[e] Specificity is the percentage of ineligible citations, as discarded by the experts, that were also excluded by the crowd member.

Individual Crowd Member's Performance

In addition to crowd performance, we evaluated individual crowd member performance as an exploratory outcome in those users having completed a minimum of 50 citation assessments (Table 4). At the abstract level, these crowd members completed a median of 306.5 assessments (IQR 108.5-513.5) and performed 98.8% (16,789/16,988) of the assessments. Individual crowd member sensitivity was calculated as a median of 96.6% (IQR 92.0-100.0), with median specificity determined to be 76.4% (IQR 66.2-92.8). At the full text level, these crowd members completed a median of 141 assessments (IQR 72-206) and performed 93.0% (7071/7604) of the assessments. Individual crowd member sensitivity was calculated as a median of 96.7% (IQR 89.6-99.0), with median specificity determined to be 64.3% (IQR 58.5-3.8). Including crowd members who have completed less than the required minimum did not have any substantial differences on the results (Multimedia Appendix 8). Individual crowd member's performance separated by SR is presented separately (Multimedia Appendix 9).

Discussion

Summary of Results

This study focused on crowdsourcing the citation review process and provides evidence suggesting both the feasibility of and the validity of this approach. First, using citations from 6 different SRs, we were able to establish that an online crowd was willing to assist with abstract screening, full text retrieval, and full text review. Importantly, this work also demonstrated that the online crowd showed a preference for certain reviews, with some reviews requiring incentives to attract crowd members and facilitate completion of abstract and full text screening. Second, through a comparison with the assessments performed by expert reviewers, we were able to demonstrate that the crowd had excellent sensitivity and performed more than 70% of abstract and full-text screening, depending on the threshold used for exclusion.

Feasibility of Crowd Screening Systematic Reviews

Multiple health and science initiatives have recently proven that online individuals are willing and motivated to participate in crowdsourcing projects. In addition to Wikipedia, a well-known crowdsourcing initiative, FoldIt, is an excellent example, where over 57,000 individuals have participated in an online game working to predict protein 3D structures [19], outperforming both computational and experimental methods [27]. Other examples from the medical field have shown the crowd to be able to assess images of optic disks and diagnose diabetic retinopathy [28-30]. Similarly, in this study, we were able to recruit an online crowd of volunteers that was sufficiently sized to surpass the target for abstract screening, locate 83% of full text articles, and complete the full text assessment for 5 of the 6 reviews. This finding, when combined with the observations that the time to task completion was significantly different between reviews, with certain reviews requiring gift card incentives, suggests that feasibility may be specific to the crowd—review dyad. A crowd's capacity to retrieve full text articles and screen them has not been evaluated previously, with related reports focusing solely on abstract screening. Although comparable published literature is limited, our results are consistent with other studies reporting that people are willing to perform SR tasks as either volunteers [23,24] or in exchange for payment [14,22]. For example, in a study by Mortensen et al [22], individuals working on the Amazon mTurk platform were paid to screen more than 1000 abstracts against the eligibility criteria for 4 different SRs of similar sizes, with the reviews completed in 5 to 17 days. The Cochrane Crowd initiative offers another great example supporting the feasibility of crowdsourcing [24,31]. Although published data are minimal on the Cochrane initiative, they have successfully organized an online community with thousands of individuals who have voluntarily screened over a million abstracts to identify those representing randomized controlled trials (RCTs) on humans. More recently, they have also evaluated having the crowd assist with individual Cochrane reviews, with online reports and abstracts demonstrating rapid completion of abstract screening (<5 days) [32]. Although review completion took slightly longer in our study because of the need to recruit a crowd de novo, the work from Amazon mTurk [14,22] and Cochrane [24] suggest even greater feasibility (ie, shorter times to review completion), given the immediate access to a large and sufficiently motivated crowd.

Self-reported information on training and research collected at participant sign-up in our study demonstrated that the majority were undergraduates or medical students with limited research experience. Although this approach does not allow us to comment on the performance of a more general online population, it is consistent with what has been observed and accepted by other successful crowdsourcing efforts in medicine and science. For example, surveys of crowdsourcing platforms such as Amazon mTurk and CrowdFlower have shown that crowd members are well educated, with around two-thirds having a college or advanced degree and a third being current students [33]. Furthermore, preliminary results from the Cochrane crowd suggest that more than 50% of crowd members worked in health-related fields [34]. On the basis of our results and those by Mortensen, it would now be reasonable to consider a study focused on the performance of the much large group of online workers without scientific or health training.

Crowd Performance—Sensitivity

Although the ability to recruit an online crowd willing to perform SR tasks was the essential first step, it is of equal importance to understand crowd performance. Similar to studies evaluating other alternative methodologies with the potential to facilitate citation screening, we selected sensitivity as the most important performance outcome [35]. Although no consensus study (eg, delphi, survey) has defined the minimum acceptable sensitivity, 95% has become the industry standard in the field of automated text recognition research based on original studies by Cohen [11,36]. In our work, both of the a priori algorithms not only achieved sensitivities above 95% but the sample size also allowed us to exclude 95% from the 95% lower CI. When further reducing the crowd threshold to a majority (≥50% exclusions) to allow for comparison with the Mortensen study, the sensitivity fell only marginally to 98.9% because of the crowd exclusion of 2 studies. Inspection of these 2 publications identified that each abstract presented information on 2 different studies packaged into 1 manuscript, with the first study described not meeting eligibility criteria [37,38]. Our study is the first to assess the crowd's ability to screen citations at the full text level. Similar to what is done currently by investigative teams, all articles that were retained at the abstract level at each of the 3 exclusion thresholds were moved to full text screening. The crowd showed high sensitivity at this level and did not miss any further eligible citations, even when only a simple majority was sufficient to exclude. In Mortensen's study [22], the crowd's sensitivity at the abstract level was compared against the gold standard approach of expert reviewers. Eligibility criteria were modified and slightly broadened for simplification purposes. The crowd sensitivity was lower than that of our study and varied between 86 to 93% using a threshold comparable with our 75%, and 71% to 89% when a simple majority was allowed to exclude an article.

Crowd Performance—Work Performed

Although establishing high sensitivity is essential, crowdsourcing is only valuable if it effectively decreases the work required of the investigative team. In this study, our 2 a priori defined algorithms (100% and 75% exclusion thresholds) reduced the work required by the investigative team by approximately 45% to 60% after the abstract level. Allowing the crowd to screen the full text for citations retained at the abstract level significantly increased the work performed on behalf of the investigative team (70%). This additional 10% to 25% increase in work performed would translate to between 200 and 500 fewer full text articles to screen in an SR of 2000 citations. The crowdsourcing validation study by Mortensen presented gain (specificity) as a measure of work performed [22]. Their algorithm requiring 100% agreement to exclude a citation achieved gains between 68% and 87% across the 4 SRs and saved 90% of the cost of the gold standard experts' approach. Using this definition, our gain was comparable and measured between 50% and 75% at the abstract level and further increased significantly to 75% to 85% after full text and depending on the exclusion threshold. It is important to note that the crowd has achieved high efficiency in both of these studies, despite slightly broadening the eligibility criteria.

Performance of Individual Crowd Members

As an exploratory objective, this study also sought to understand the performance of individual uncurated volunteer crowd members, with the results suggesting the average participant user to have excellent sensitivity (96%) and good specificity (70%). These findings are important as they suggest it may be possible to retain excellent project-level sensitivity with fewer crowd assessments per citations. Reducing the number of assessments per citation could have multiple advantages, including reducing the time to individual project completion, increasing the number of projects a crowd of set size can assist with, and maximizing work performed (specificity). Although most crowd members performed well, 4% of the crowd were observed to have less than adequate sensitivities (<80%). Although it only represents a minority, inclusion of 1 or more of these poorly performing crowd members could place a project at risk if the number of assessments per citation was significantly reduced. Although the goal in our study was to evaluate the performance of an uncurated crowd, we acknowledge that it would have been reasonable and potentially beneficial to attempt the removal of these poorly performing individuals by requiring the successful completion of a test set. This approach was employed in the 2 crowdsourcing studies utilizing the Amazon Mechanical Turk system, where workers were required to successfully evaluate 3 articles before being invited to the full project [14,22]. Although this approach would have had benefits, the authors did also observe that initial testing alone was insufficient as some reviewers developed "unconscientious" behavior over time that required embedded quality control or "honey pots" [22]. Available evidence suggests that with the right combination of initial testing and ongoing monitoring, it will be possible to further optimize crowd sensitivity and work performed [39,40]. It will be important for future studies to establish the initial and embedded testing required to guarantee comprehensive SRs, while optimizing crowd work. The size and components of the test set, the threshold for sensitivity, and how to embed quality control will need to be evaluated as part of larger studies.

Crowdsourcing—Barriers to Implementation and Future Directions

Although crowdsourcing has the potential to lead to more rapid knowledge synthesis and evidence translation, it is important to acknowledge that it can only do so if accessible, cost-effective, and scalable. Presently, and similar to what transpires in other areas of interventional and diagnostic research, the innovation is initially only available to a handful of individual teams and organizations who have taken the considerable time to both develop a platform (Cochrane Crowd, CrowdscreenSR) or adopt one (Amazon mTurk) and recruit a crowd. Although it may not be possible for other SR teams requiring a crowd for a large project to access the exact individual or crowds utilized in the existing feasibility and validation studies, overall findings do suggest it would be possible to rapidly recruit a similarly sized and motivated crowd through emails and promotional materials. Consider, for example, that each major center in North America, and beyond, has hundreds potentially thousands of undergraduate medical students, residents, and health care professionals who may want to engage in knowledge synthesis efforts. Although some large institutions and organization, similar to Cochrane, may consider creating their own SR crowdsourcing initiative, there are considerable costs associated with the development and maintenance of a user-friendly robust platform that allows investigators to present projects and both evaluate and track crowd performance through the citation review process. Consequently, the ideal future state includes the development of an online SR citation screening platform broadly available to a wide range of institutions, organization, and countries that share both the costs and benefits of the platform. As the success of such an initiative would necessitate engaging with a large online distributed crowd with a broad range of interest and experiences, future work in this area should seek to understand what motivates individuals to assist with crowdsourcing SR tasks [41]. It is worth noting that the aforementioned SR crowdsourcing studies and initiatives have been able to succeed using motivators such as certain types of payment and volunteer or research experience. Whether these would be sufficient on a large scale remains to be determined. Missing from this list are more objective measures of academic credit, including group or named authorship. Although not part of our original study protocol, many of the individual crowd members expressed interest in, and have since participated in, SRs for named authorship or as part of a group at our institute [42].

Crowdsourcing and Machine Learning

Future work aimed at developing a platform capable of facilitating and optimizing crowdsourcing into SR should also consider incorporating automated or computerized abstract screening. This has been hypothesized and investigated as an alternative means of reducing the work required by SR investigative teams. The findings in our crowdsourcing study and those reported by Mortensen [35] are similar to or exceed the 30% to 70% reported in text-mining studies. Machine learning has shown strong accuracy and cost-effectiveness when studies have focused on a single screening criterion (ie, study design—RCT or not). Where multiple elements of the articles need to be assessed, machine learning can require considerable costs related to training. Although a comparison of crowdsourcing with text-mining performance is valid, it is also worthwhile considering that by combining machine learning and crowdsourcing together may lead to the greatest workload reduction for the crowd and investigative teams [43,44]. This hybrid approach has been researched and applied in a variety of fields outside the SR field. As an immediately relatable example, Google employs machine learning to generate search results, which are then further improved by integrating users' selection [45]. For SR screening, this combined approach would involve having the machine learn on an initial training set prepared by the investigative team, followed by identification of very low probability citations, using machine learning, with the remaining referred to the crowd. The lone published study to consider this approach by Wallace et al had a machine-learning algorithm to identify citations unlikely to be an RCT. This approach eliminated 80% of the citations, with the remaining 20% containing 98% of the eligible citations [43]. Interestingly, the authors estimated that this approach could reduce study costs by 7-fold. Another approach proposed by Bannach-Brown would be to use crowd's assessments on a training set to develop the machine-learning approach, which would be later reused on the training set (to identify potential errors) or the remaining of citations [46].

Study Strengths and Limitations

This study offers a significant contribution to the emerging field of SR crowdsourcing. It is the first to report on crowd members' demographics, their capacity for full text retrieval, and performance on evaluating full text. However, certain limitations must be acknowledged. First, although our study provides evidence supporting the feasibility of crowdsourcing, the platform used for the study and the exact crowd are no longer available. Fortunately, recruitment of a sufficiently sized crowd (>30) allows us to provide a reliable estimate of the performance of the much larger population of individuals with science or health training distributed across centers and cities around the world who might consider participating in SR projects. Second, demographic information collected on the crowd participants determined that the majority were from the Ottawa region and had some postsecondary education. As demographic and training data were self-reported, we cannot be certain of the accuracy of these data and the implications of misclassification (eg, falsely elevating experience). Given the uncertainty about both the accuracy of self-reported education and training data and generalizability of the results to more geographically diverse crowds, we would recommend that scientists and clinicians incorporate initial and embedded quality control measures in an SR crowdsourcing initiative. Third, although our study suggests that it may be possible, perhaps beneficial, to consider fewer crowd assessments per citation, in the setting of initial and ongoing testing, it is not yet possible to provide definitive guidance. Fourth, although this study evaluated crowd performance on citations from 6 different SRs, the largest to date, it is not yet clear how well the results extrapolate to reviews from divergent areas. Our findings would generalize best to SR focused on health and those focused on children. Finally, this study was not properly designed to evaluate or comment on cost or time saving. As an early exploratory pilot work that required the development of a software platform and crowd recruitment, it is likely that no benefit would have been observed. We would suggest that this work be reserved for after the development of the aforementioned user-friendly robust online platform and recruitment of a sufficiently sized motivated crowd.

Conclusions

This study supports the feasibility and validity of crowdsourcing as a means to facilitate citation screening for SRs by a crowd of nonexpert volunteers with some medical and/or scientific background. It also offers the first evidence for screening at the full text level. This approach is not intended to replace the gold standard expert screening but rather to supplement it by expediting the screening process, thus allowing the investigative team to focus on more complex SR tasks. To get the full potential benefits of crowdsourcing, future projects should aim at establishing a comparable platform that would allow researchers to easily access a large and expanding crowd similar to the one recruited here. Future directions should assess the motivation of the crowd, what incentives could improve performance, how to predict the crowd members with higher performance, and the need for quality control measures such as honeypots.

Acknowledgments

This study was funded by a CHEO Research Institute Research Growth Award. CHEO Research Institute had no role in the design and conduct of the study, analysis of the data, and preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Privacy policy and terms of use.

[PDF File (Adobe PDF File), 116KB - jmir_v21i4e12953_app1.pdf]

Multimedia Appendix 2

Standards for Reporting of Diagnostic Accuracy Studies guidelines.

[PDF File (Adobe PDF File), 715KB - jmir_v21i4e12953_app2.pdf]

Multimedia Appendix 3

Promotional material.

[PDF File (Adobe PDF File), 7MB - jmir_v21i4e12953_app3.pdf]

Multimedia Appendix 4

Citations disposition based on crowd’s assessment using multiple exclusion thresholds.

[PDF File (Adobe PDF File), 40KB - jmir_v21i4e12953_app4.pdf]

Multimedia Appendix 5

Number of assessments per paper at each screening level.

[PDF File (Adobe PDF File), 52KB - jmir_v21i4e12953_app5.pdf]

Multimedia Appendix 6

Crowd’s sensitivity and work performed by systematic review at different exclusion thresholds for the abstract level.

[PDF File (Adobe PDF File), 53KB - jmir_v21i4e12953_app6.pdf]

Multimedia Appendix 7

Crowd’s sensitivity and work performed by systematic review at different exclusion thresholds.

[PDF File (Adobe PDF File), 53KB - jmir_v21i4e12953_app7.pdf]

Multimedia Appendix 8

Individual crowd members’ performance including those who completed less than 50 assessments.

[PDF File (Adobe PDF File), 51KB - jmir_v21i4e12953_app8.pdf]

Multimedia Appendix 9

Individual crowd members’ performance by systematic review.

[PDF File (Adobe PDF File), 53KB - jmir_v21i4e12953_app9.pdf]

References

1. Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. Int J Evid Based Healthc 2005 Sep;3(8):207-215. [doi: 10.1111/j.1479-6988.2005.00026.x] [Medline: 21631749]
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
3. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 2007 Aug 21;147(4):224-233. [doi: 10.7326/0003-4819-147-4-200708210-00179] [Medline: 17638714]
4. Xu X, Li Y, Sheng Y, Liu J, Tang L, Chen Z. Effect of low birth weight on childhood asthma: a meta-analysis. BMC Pediatr 2014 Oct 23;14:275 [FREE Full text] [doi: 10.1186/1471-2431-14-275] [Medline: 25339063]
5. Kelley GA, Kelley KS, Pate RR. Effects of exercise on BMI z-score in overweight and obese children and adolescents: a systematic review with meta-analysis. BMC Pediatr 2014 Sep 9;14:225 [FREE Full text] [doi: 10.1186/1471-2431-14-225] [Medline: 25204857]
6. Sun J, Zheng J, Tang L, Healy J, Sinnett D, Dai Y. Association between CEBPE variant and childhood acute leukemia risk: evidence from a meta-analysis of 22 studies. PLoS One 2015;10(5):e0125657 [FREE Full text] [doi: 10.1371/journal.pone.0125657] [Medline: 25938438]
7. Wallace AE, Neily J, Weeks WB, Friedman MJ. A cumulative meta-analysis of selective serotonin reuptake inhibitors in pediatric depression: did unpublished studies influence the efficacy/safety debate? J Child Adolesc Psychopharmacol 2006;16(1-2):37-58. [doi: 10.1089/cap.2006.16.37] [Medline: 16553528]
8. Allen IE, Olkin I. Estimating time to conduct a meta-analysis from number of citations retrieved. J Am Med Assoc 1999 Aug 18;282(7):634-635. [doi: 10.1001/jama.282.7.634] [Medline: 10517715]
9. Jonnalagadda S, Petitti D. A new iterative method to reduce workload in systematic review process. Int J Comput Biol Drug Des 2013;6(1-2):5-17 [FREE Full text] [doi: 10.1504/IJCBDD.2013.052198] [Medline: 23428470]
10. Matwin S, Kouznetsov A, Inkpen D, Frunza O, O'Blenis P. A new algorithm for reducing the workload of experts in performing systematic reviews. J Am Med Inform Assoc 2010;17(4):446-453 [FREE Full text] [doi: 10.1136/jamia.2010.004325] [Medline: 20595313]
11. Cohen AM, Hersh WR, Peterson K, Yen P. Reducing workload in systematic review preparation using automated citation classification. J Am Med Inform Assoc 2006;13(2):206-219 [FREE Full text] [doi: 10.1197/jamia.M1929] [Medline: 16357352]
12. Créquit P, Trinquart L, Yavchitz A, Ravaud P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. BMC Med 2016 Jan 20;14:8 [FREE Full text] [doi: 10.1186/s12916-016-0555-0] [Medline: 26792360]
13. Nama N, Menon K, Iliriani K, Pojsupap S, Sampson M, O'Hearn K, et al. A systematic review of pediatric clinical trials of high dose vitamin D. PeerJ 2016;4:e1701 [FREE Full text] [doi: 10.7717/peerj.1701] [Medline: 26966655]
14. Brown AW, Allison DB. Using crowdsourcing to evaluate published scientific literature: methods and example. PLoS One 2014;9(7):e100647 [FREE Full text] [doi: 10.1371/journal.pone.0100647] [Medline: 24988466]
15. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203 [FREE Full text] [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]
16. Lee YJ, Arida JA, Donovan HS. The application of crowdsourcing approaches to cancer research: a systematic review. Cancer Med 2017 Nov;6(11):2595-2605 [FREE Full text] [doi: 10.1002/cam4.1165] [Medline: 28960834]
17. Rossiter DG, Liu J, Carlisle S, Zhu A. Can citizen science assist digital soil mapping? Geoderma 2015 Dec;259-260:71-80. [doi: 10.1016/j.geoderma.2015.05.006]
18. Lintott C, Schawinski K, Slosar A, Land K, Bamford S, Thomas D, et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Mon Not R Astron Soc 2008 Sep 21;389(3):1179-1189. [doi: 10.1111/j.1365-2966.2008.13689.x]
19. Good BM, Su AI. Games with a scientific purpose. Genome Biol 2011 Dec 28;12(12):135. [doi: 10.1186/gb-2011-12-12-135] [Medline: 22204700]

20. Vachelard J, Gambarra-Soares T, Augustini G, Riul P, Maracaja-Coutinho V. A guide to scientific crowdfunding. PLoS Biol 2016 Feb;14(2):e1002373 [FREE Full text] [doi: 10.1371/journal.pbio.1002373] [Medline: 26886064]
21. Candido DR, Lynn S, Ali HR, Eccles D, Hanby A, Provenzano E, et al. Crowdsourcing the general public for large scale molecular pathology studies in cancer. EBioMedicine 2015 Jul;2(7):681-689 [FREE Full text] [doi: 10.1016/j.ebiom.2015.05.009] [Medline: 26288840]
22. Mortensen ML, Adam GP, Trikalinos TA, Kraska T, Wallace BC. An exploration of crowdsourcing citation screening for systematic reviews. Res Synth Methods 2017 Sep;8(3):366-386 [FREE Full text] [doi: 10.1002/jrsm.1252] [Medline: 28677322]
23. Nama N, Iliriani K, Xia MY, Chen BP, Zhou LL, Pojsupap S, et al. A pilot validation study of crowdsourcing systematic reviews: update of a searchable database of pediatric clinical trials of high-dose vitamin D. Transl Pediatr 2017 Jan;6(1):18-26 [FREE Full text] [doi: 10.21037/tp.2016.12.01] [Medline: 28164026]
24. Elliott J, Sim I, Thomas J, Owens N, Dooley G, Riis J, et al. #CochraneTech: technology and the future of systematic reviews. Cochrane Database Syst Rev 2014(9):ED000091. [doi: 10.1002/14651858.ED000091] [Medline: 25288182]
25. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Br Med J 2015 Oct 28;351:h5527 [FREE Full text] [doi: 10.1136/bmj.h5527] [Medline: 26511519]
26. Kantor N, Wayne C, Nasr A. Symptom development in originally asymptomatic CPAM diagnosed prenatally: a systematic review. Pediatr Surg Int 2018 Jun;34(6):613-620. [doi: 10.1007/s00383-018-4264-y] [Medline: 29632964]
27. Khatib F, DiMaio F, Foldit Contenders Group, Foldit Void Crushers Group, Cooper S, Kazmierczyk M, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat Struct Mol Biol 2011 Sep 18;18(10):1175-1177 [FREE Full text] [doi: 10.1038/nsmb.2119] [Medline: 21926992]
28. Mitry D, Peto T, Hayat S, Blows P, Morgan J, Khaw K, et al. Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. PLoS One 2015;10(2):e0117401 [FREE Full text] [doi: 10.1371/journal.pone.0117401] [Medline: 25692287]
29. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res 2014;16(10):e233 [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]
30. Wang X, Mudie L, Brady CJ. Crowdsourcing: an overview and applications to ophthalmology. Curr Opin Ophthalmol 2016 May;27(3):256-261 [FREE Full text] [doi: 10.1097/ICU.0000000000000251] [Medline: 26761188]
31. Tsertsvadze A, Chen Y, Moher D, Sutcliffe P, McCarthy N. How to conduct systematic reviews more expeditiously? Syst Rev 2015 Nov 12;4:160 [FREE Full text] [doi: 10.1186/s13643-015-0147-7] [Medline: 26563648]
32. Featherstone R, McDonald S, Noel-Storr A, Thomas J. Advanced Information Retrieval on the Edge Living Systematic Reviews. 2018 Presented at: Health Technology Assessment international; June 1-5, 2018; Vancouver, Canada.
33. Ross J, Irani L, Silberman M, Zaldivar A, Tomlinson B. Who are the Crowdworkers? Shifting Demographics in Mechanical Turk. In: Extended Abstracts on Human Factors in Computing Systems. 2010 Apr 10 Presented at: CHI EA'10 CHI'10; April 10-15, 2010; Atlanta, GA p. 2863-2872. [doi: 10.1145/1753846.1753873]
34. Noel-Storr A, Thomas J, Mavergames C, Turner T, McDonald S, Green S, et al. Cochrane Crowd: using citizen science to meet the challenge of information overload in evidence production. 2016 Oct 23 Presented at: 24th Cochrane Colloquium; October 23-26, 2016; Seoul, Korea. [doi: 10.13140/RG.2.2.11290.34240]
35. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 2015 Jan 14;4:5 [FREE Full text] [doi: 10.1186/2046-4053-4-5] [Medline: 25588314]
36. Cohen AM. Optimizing feature representation for automated systematic review work prioritization. AMIA Annu Symp Proc 2008 Nov 6:121-125 [FREE Full text] [Medline: 18998798]
37. Com G, Carroll JL, Tang X, Melguizo MS, Bower C, Jambhekar S. Characteristics and surgical and clinical outcomes of severely obese children with obstructive sleep apnea. J Clin Sleep Med 2015 Apr 15;11(4):467-474 [FREE Full text] [doi: 10.5664/jcsm.4608] [Medline: 25665695]
38. Roche SL, Timberlake K, Manlhiot C, Balasingam M, Wilson J, George K, et al. Angiotensin-converting enzyme inhibitor initiation and dose uptitration in children with cardiovascular disease: a retrospective review of standard clinical practice and a prospective randomized clinical trial. J Am Heart Assoc 2016 Dec 20;5(5):e003230 [FREE Full text] [doi: 10.1161/JAHA.116.003230] [Medline: 27207965]
39. Simperl E. How to use crowdsourcing effectively: guidelines and examples. LIBER Q 2015 Aug 18;25(1):18. [doi: 10.18352/lq.9948]
40. Raykar V, Yu S. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. J Mach Learn Res 2012 Feb;13:491-518 [FREE Full text] [doi: 10.1162/15324430152748236]
41. Sarti AJ, Sutherland S, Landriault A, DesRosier K, Brien S, Cardinal P. Exploring the components of physician volunteer engagement: a qualitative investigation of a national Canadian simulation-based training programme. BMJ Open 2017 Dec 23;7(6):e014303 [FREE Full text] [doi: 10.1136/bmjopen-2016-014303] [Medline: 28645956]

42. Kuenzig ME, Bishay K, Leigh R, Kaplan GG, Benchimol EI, Crowdscreen SR Review Team. Co-occurrence of asthma and the inflammatory bowel diseases: a systematic review and meta-analysis. Clin Transl Gastroenterol 2018 Sep 24;9(9):188 [FREE Full text] [doi: 10.1038/s41424-018-0054-z] [Medline: 30250122]
43. Wallace B, Noel-Storr A, Marshall I, Cohen A, Smalheiser N, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. J Am Med Inform Assoc 2017 Nov 1;24(6):1165-1168 [FREE Full text] [doi: 10.1093/jamia/ocx053] [Medline: 28541493]
44. Krivosheev E, Casati F, Baez M, Benatallah B. Combining crowd and machines for multi-predicate item screening. Proc ACM Hum Comput Interact 2018 Nov 1;2(CSCW):1-18. [doi: 10.1145/3274366]
45. Vaughan J. Making better use of the crowd: how crowdsourcing can advance machine learning research. J Mach Learn Res 2018 May;18(193):46 [FREE Full text]
46. Bannach-Brown A, Przybyła P, Thomas J, Rice A, Ananiadou S, Liao J, et al. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Syst Rev 2019 Jan 15;8(1):23 [FREE Full text] [doi: 10.1186/s13643-019-0942-7] [Medline: 30646959]

Abbreviations

3D: three-dimensional
CHEO: Children’s Hospital of Eastern Ontario
IQR: interquartile range
RCT: randomized controlled trial
SR: systematic review

Edited by G Eysenbach; submitted 23.12.18; peer-reviewed by B Wallace, H Zheng, C Basch; comments to author 24.01.19; revised version received 18.02.19; accepted 24.03.19; published 29.04.19.

Please cite as:
Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G, Murto K, Vaccani JP, Katz S, Zemek R, Nasr A, McNally JD
Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study
J Med Internet Res 2019;21(4):e12953
URL: http://www.jmir.org/2019/4/e12953/
doi:10.2196/12953
PMID:31033444

Review

Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature

Andreas K Triantafyllidis[1,2], BSc, MSc, PhD; Athanasios Tsanas[3,4], BSc, BEng, MSc, PhD

[1]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

[2]Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[3]Usher Institute of Population Health Sciences and Informatics, Medical School, University of Edinburgh, Edinburgh, United Kingdom

[4]Mathematical Institute, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Andreas K Triantafyllidis, BSc, MSc, PhD
Information Technologies Institute
Centre for Research and Technology Hellas
6th km Charilaou-Thermi Rd
GR 57001 Thermi
Thessaloniki, 60361
Greece
Phone: 30 2310 498100
Fax: 30 2310 498180
Email: atriand@gmail.com

Abstract

Background: Machine learning has attracted considerable research interest toward developing smart digital health interventions. These interventions have the potential to revolutionize health care and lead to substantial outcomes for patients and medical professionals.

Objective: Our objective was to review the literature on applications of machine learning in real-life digital health interventions, aiming to improve the understanding of researchers, clinicians, engineers, and policy makers in developing robust and impactful data-driven interventions in the health care domain.

Methods: We searched the PubMed and Scopus bibliographic databases with terms related to machine learning, to identify real-life studies of digital health interventions incorporating machine learning algorithms. We grouped those interventions according to their target (ie, target condition), study design, number of enrolled participants, follow-up duration, primary outcome and whether this had been statistically significant, machine learning algorithms used in the intervention, and outcome of the algorithms (eg, prediction).

Results: Our literature search identified 8 interventions incorporating machine learning in a real-life research setting, of which 3 (37%) were evaluated in a randomized controlled trial and 5 (63%) in a pilot or experimental single-group study. The interventions targeted depression prediction and management, speech recognition for people with speech disabilities, self-efficacy for weight loss, detection of changes in biopsychosocial condition of patients with multiple morbidity, stress management, treatment of phantom limb pain, smoking cessation, and personalized nutrition based on glycemic response. The average number of enrolled participants in the studies was 71 (range 8-214), and the average follow-up study duration was 69 days (range 3-180). Of the 8 interventions, 6 (75%) showed statistical significance (at the P=.05 level) in health outcomes.

Conclusions: This review found that digital health interventions incorporating machine learning algorithms in real-life studies can be useful and effective. Given the low number of studies identified in this review and that they did not follow a rigorous machine learning evaluation methodology, we urge the research community to conduct further studies in intervention settings following evaluation principles and demonstrating the potential of machine learning in clinical practice.

(J Med Internet Res 2019;21(4):e12286) doi:10.2196/12286

KEYWORDS

machine learning; data mining; artificial intelligence; digital health; review; telemedicine

Introduction

Background

Digital health interventions [1], including modalities such as telemedicine, Web-based strategies, email, mobile phones, mobile apps, text messaging, and monitoring sensors, have enormous potential to support independent living and self-management [2], and reduce health care costs [3]. They have also shown great promise in improving health [4]. With the advent of new tools and algorithms for machine learning, a new class of smart digital health interventions can be developed, which could revolutionize effective health care delivery [5].

The term *machine learning* is widely used across disciplines but has no universally accepted definition [6]. This is in part explained by the breadth of the areas it covers and because researchers from diverse disciplines have historically contributed (and still contribute) to its development. Broadly, it refers to an algorithmic framework that can provide insights into data, while facilitating inference and providing a tentative setting to determine functional relationships.

Machine learning has been applied in multiple health care domains, including diabetes [7], cancer [8], cardiology [9], and mental health [10]. Most of the developed machine learning models and tools in research settings have investigated the potential of prognosis [11], diagnosis [12], or differentiation of clinical groups (eg, a group with a pathology and a healthy control group or groups with pathologies) [13], thus demonstrating promise toward the development of computerized decision support tools [14]. The key requirements for the development of these tools are sufficiently large datasets (in terms of both number of participants and explanatory variables to explore) and accurate labels, typically provided by expert clinicians. The premise is the identification of those data structures or variables (eg, clinical, behavioral, or demographic variables) that are associated with the target outcome (eg, whether a person has cancer). In this regard, useful knowledge can be derived from the available data, which can empower patients to monitor their health status longitudinally and support health professionals in decision making with regard to management, treatment, and follow-up interventions where required.

Despite a considerably growing body of research literature in the use of machine learning in health care applications [15], it is astonishing how few of these suggestions are actually translated into clinical practice [16]. There is remarkably limited empirical evidence of the effectiveness of machine learning applications in digital health interventions. This is rather surprising, since any proposed health care solutions would reach their full potential only if they are embraced by the medical community, becoming integrated within properly designed digital health interventions and tested in real-life studies with patients and health professionals.

Objective

Considering that machine learning models and tools have not been widely and reliably used in clinical practice, whereas the peer-reviewed literature in the field is growing exponentially, we wanted to assess the progress made in smart data-driven health interventions applied in real-life research settings—that is, the real world in which constraints in available resources or opportunities to collect reliable data may exist, as opposed to simulation or laboratory-based studies [17]. In this direction, we present a systematic literature review of digital health interventions incorporating machine learning algorithms, by identifying and mapping their features and outcomes, with the aim to improve our knowledge of the design and development of impactful intelligent interventions.

Methods

Inclusion and Exclusion Criteria

We sought to identify digital health nonpharmacological interventions incorporating machine learning that were assessed in pragmatic studies. In this context, the inclusion criteria for study selection were (1) the study should be conducted with patients or health professionals, or both, in a real-life setting, (2) machine learning algorithms or models were used in the digital health intervention (rather than merely reporting statistical hypothesis testing results or statistical associations), (3) quantitative outcomes of the study were presented, and (4) the article describing the study was written in English. We excluded retrospective studies, case reports, ongoing studies, surveys or reviews, laboratory or simulation studies, studies describing protocols, qualitative studies, and all studies published before 2008 from the review because we wanted to determine the status of recent research developments in the field that have been used in clinical interventional settings.

Literature Search and Screening

We searched the PubMed and Scopus bibliographic databases for studies published after 2008 using the string "(machine learning) OR (data mining) OR (artificial intelligence) AND health" for search within the title, abstract, and keywords of the articles. We limited "Species" in PubMed to humans.

Both authors independently screened the identified articles following the literature search to minimize bias in the selection process. Any disagreements were resolved by discussion between the authors and reaching a consensus. We screened the abstracts of the candidate articles for inclusion and subsequently read the full text of the articles deemed eligible according to the inclusion criteria. Subsequently, we excluded articles not providing sufficient information about the application of machine learning or for being ineligible. We used the Effective Public Health Practice Project (EPHPP) tool to assess the methodological quality of the included studies, which has been found to be reliable [18]. The studies that focused on interventions were synthesized (AKT) according to their target (ie, target condition), study design, number of enrolled participants, follow-up duration, primary outcome and whether this was significantly positive, machine learning algorithms used in the intervention, and outcome of the algorithms (eg, prediction of a target outcome).

The systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses

(PRISMA) guidelines [19]. Multimedia Appendix 1 shows a completed PRISMA checklist.

Results

Literature Search Outcomes

Our last search in November 2018 returned 1386 articles from the PubMed database and 7024 articles from Scopus. We imported all the retrieved records into Mendeley (version 1.19.3) bibliography management software (Mendeley Ltd) [20], which identified 1093 duplicates. We screened the abstracts of the remaining 7317 results according to our inclusion and exclusion criteria and identified 21 eligible articles. The reviewers read the full text of the 21 articles and agreed on 8 for inclusion as eligible articles. The flow diagram in Figure 1 summarizes the reasons for excluding research articles for study inclusion following the PRISMA format (Figure 1).

Quality Assessment

On the basis of the EPHPP criteria for selection bias, design, confounders, blinding, data collection, and dropouts, we found the methodological quality to be moderate for 2 of the 8 (25%) studies [21,22] and weak for the remaining 6 (75%) studies [23-28] (Table 1). Most studies were poorly rated because of selection bias, insufficient care in controlling for confounders, and the high percentage of withdrawals or dropouts (or the absence of their description). The design of a randomized or controlled clinical trial was described in 3 (37%) studies [21,25,28], and 5 (63%) interventions were evaluated in a pilot or experimental single-group study (Multimedia Appendix 2).

Figure 1. Flow diagram for study inclusion following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) format.

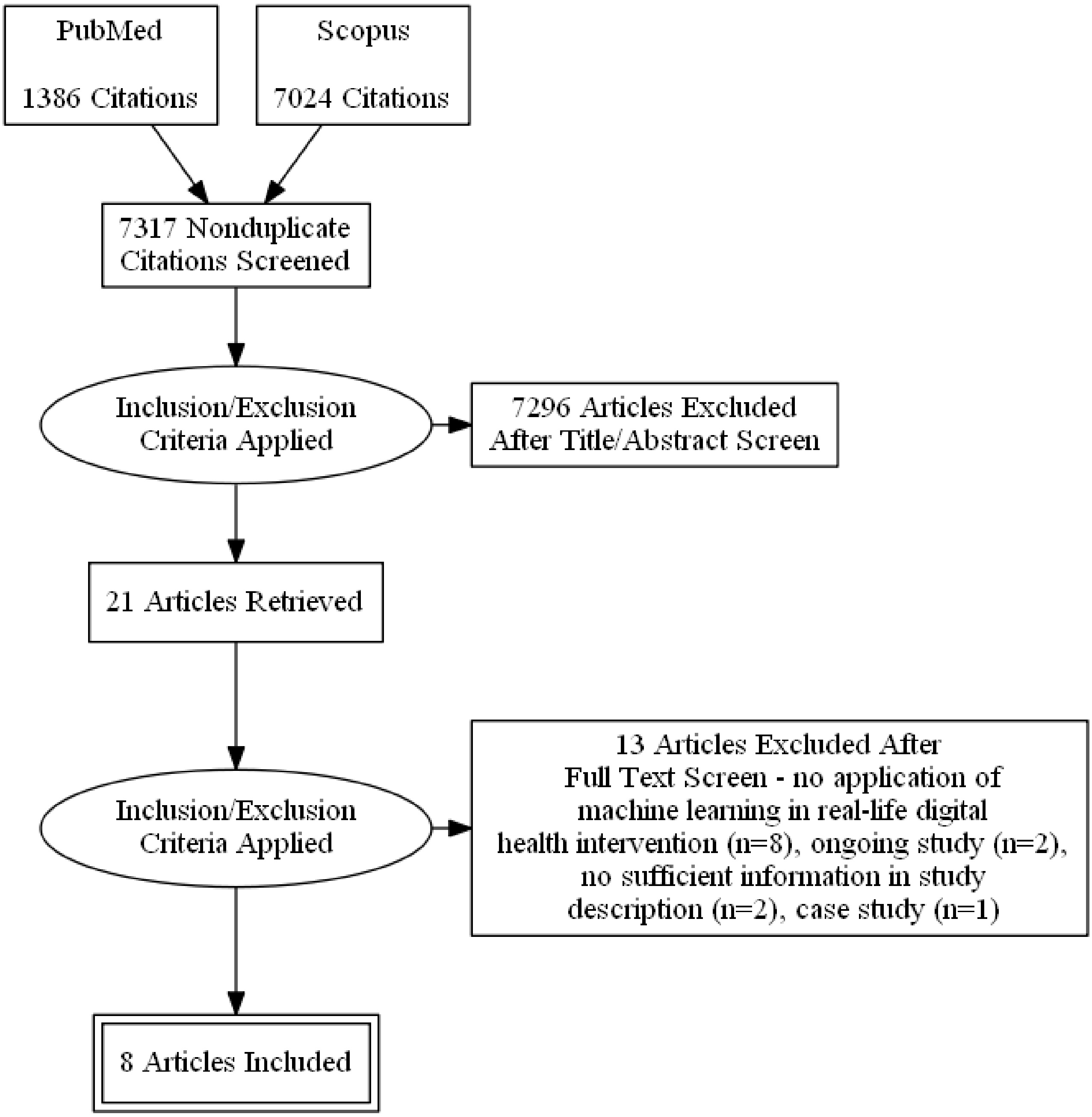


Table 1. Quality assessment of included studies based on Effective Public Health Practice Project (EPHPP) criteria.

Study first author, year, and reference	EPHPP criteria						Global rating
	Selection bias	Study design	Confounders	Blinding	Data collection methods	Withdrawals and dropouts	
Burns, 2011 [23]	Weak	Moderate	Weak	Moderate	Strong	Strong	Weak
Hawley, 2013 [24]	Weak	Weak	Weak	Moderate	Weak	Weak	Weak
Manuvinakurike, 2014 [22]	Weak	Strong	Strong	Moderate	Moderate	Strong	Moderate
Martin, 2012 [25]	Moderate	Strong	Weak	Moderate	Strong	Weak	Weak
Morrison, 2017 [26]	Moderate	Moderate	Weak	Strong	Strong	Weak	Weak
Ortiz-Catalan, 2016 [27]	Moderate	Moderate	Weak	Moderate	Weak	Weak	Weak
Sadasivam, 2016 [21]	Weak	Strong	Strong	Strong	Moderate	Strong	Moderate
Zeevi, 2015 [28]	Weak	Strong	Weak	Moderate	Weak	Weak	Weak

Type of Intervention and Target Population

The interventions targeted depression prediction and management [23], speech recognition for people with speech disabilities [24], self-efficacy for weight loss [22], detection of changes in biopsychosocial condition of patients with multiple morbidity [25], stress management [26], treatment of phantom limb pain [27], smoking cessation [21], and personalized nutrition based on glycemic response [28] (Multimedia Appendix 2).

Of the 8 interventions, 3 (37%) targeted patients: individuals with a diagnosis of depression [23], those with multiple morbidities such as lung disease and cardiovascular disease [25], and those with phantom limb pain [27]. One (13%) intervention targeted people with speech disabilities [24]; 4 (50%) interventions targeted individuals who had no explicit diagnosis of a disease or impairment [21,22,26,28]. All target groups comprised adults. The average number of enrolled participants in the studies was 71 (range 8-214), and the average follow-up study duration was 69 days (range 3-180).

Applications of Machine Learning and Outcomes

Overall, 6 of the 8 (75%) real-life studies of digital health interventions aided by machine learning algorithms showed statistical significance (at the *P*=.05 level) in health outcomes. Different summary measures were used in the identified studies to assess primary outcomes, which reflects the lack of standardization both in methodology and in the metrics used in the research fields. Where possible, we aimed to use the accuracy of the algorithms used and the *P* value (eg, for showing statistical significance of outcomes in an intervention group compared with a control group) as the principal summary measures. We briefly describe all included studies below in terms of intervention purpose and content, evaluation outcomes, and implications for clinical practice.

Burns et al [23] described a multicomponent mobile-based intervention that used machine learning models to predict the mood, emotions, cognitive and motivational states, activities, and environmental and social context of patients with depression, along with feedback graphs for self-reflection on behavior and coaching provided by caregivers. The predictive models were based on phone sensor–derived variables (eg, global positioning system, ambient light, phone calls), and regression along with decision trees was used. The accuracy of the models was promising for location prediction (60%-91%), but prediction was very poor for emotions such as sadness. Overall, the 8 participants in the study became less likely to meet the criteria for a diagnosis of major depressive disorder (*P*=.03), and their symptoms of depression and anxiety were decreased by the end of the study (*P*<.001). Patients were also satisfied with the intervention (5.71 average rating on a scale 1 to 7), and 6 of 7 treatment completers (86%) indicated that the intervention was helpful in understanding triggers for negative moods. Despite the benefits of self-reflection on behavior through the use of a multicomponent mobile health monitoring system and the clinical improvements shown in the study, the authors reported that the clinical utility of the prediction models they used should be improved, since the prediction outcomes (eg, location and mood) were merely displayed to the users, and there were no direct interventions based on them.

Hawley et al [24] described the use of a device capable of recognizing the speech of people with dysarthria and generating voice messages. The authors used hidden Markov models to determine the proximity of a spoken word to a personalized speech model for that individual. However, only 67% recognition accuracy was achieved in this real-life observational study with 9 participants. Participants noticed that ease of communication was reduced through the device compared with their usual communication method of either speaking or speaking supported by a conventional voice-output communication aid, mainly due to the low accuracy of speech recognition. Nevertheless, feedback from participants was positive about the device's concept, given that speech recognition was improved.

Manuvinakurike et al [22] focused on changes in self-efficacy for weight loss through the provision of personal health behavior change stories found on the internet. An algorithm based on adaptive boosting was developed to find the most relevant story based on the stage of change and the demographic characteristics of a user, along with the emotional tone and overall quality of the story (accuracy between 84% and 98% for the classification of 5 stages of change). Testing of the algorithm with 103 users revealed significantly greater increases in self-efficacy for weight loss (*P*=.02) and a statistically insignificant effect on change in decisional balance (*P*=.83). In addition, the medium used to tell the stories, being either text or an animated conversational agent, had no effect on health behavior change. The authors concluded that their approach could maximize participants' engagement in longitudinal health behavior change interventions.

Martin et al [25] used a system in which decision trees could predict unplanned hospital visits of patients with multiple morbidities such as lung disease or cardiovascular disease. Alerts were sent to health professionals, who acted on the alerts according to agreed guidelines. The system was based on information received via patient phone calls with lay care guides. Linguistic and metalinguistic features were extracted, together with the patient's status, to train the prediction models (positive predictive value of 70% for predicting unplanned events). A randomized controlled trial with 214 patients for 6 months (the largest trial we found in the review in terms of number of enrolled participants and duration) showed a reduction of 50% in the number of unplanned hospital events of participants in the intervention group compared with control. The most common response to an alert indicating that a patient needed attention (red alert) was to phone the patient the next day to reassess the situation and contact their general practitioner (3% of calls), suggest or plan a visit to their general practitioner (11% of calls), or call an ambulance (<0.01% of calls). In summary, the authors reported that predictive analytics on an ongoing basis could be used to signify risk of hospitalization and guide the health care system to take appropriate actions.

Morrison et al [26] used push notifications to enhance engagement of smartphone users for stress management. They used a naïve Bayes classifier to predict whether a user would respond to a notification, thereby building a personalized intelligent mechanism for notification delivery, based on the times within a day a user was more likely to view and react to the received messages. However, this exploratory study with 77 participants showed no statistically significant difference between participants receiving the messages sent "intelligently" and those receiving a message daily or occasionally within 72 hours (Cohen *d*=0.14 for intelligent vs daily group and *d*=0.5 for intelligent vs occasional group, for actions taken in response to messages). Although notification delivery based on time had no effect on the study groups (ie, response to notifications was no different), the authors concluded that frequent daily messages may not deter users from engaging with digital health interventions.

Ortiz-Catalan et al [27] applied myoelectric pattern recognition algorithms for the control of a virtual limb in patients with phantom limb pain and used gaming along with augmented and virtual reality for treatment. This single-group study with 14 participants revealed that patients' symptoms of phantom limb pain were significantly decreased (by about 50%) at the end of the provided treatment for 6 months (*P*=.0001 for reduction in intensity and quality of pain). The authors suggested that their novel treatment could be used after failure of evidence-based treatments such as mirror therapy and before proceeding with invasive or pharmacological approaches.

Sadasivam et al [21] used a recommender system to send motivational messages to individuals, targeting smoking cessation. The system was based on Bayesian probabilistic matrix factorization to predict message rating, through the processing of data from the user's previous ratings of messages, along with other users' ratings. This randomized controlled trial with 120 users showed that the system was more effective at influencing people to quit smoking than were standard tailored messages (rule-based system) with proven effectiveness (*P*=.02) and resulted in a similar cessation rate. The authors concluded that their recommender system could be used instead of standard systems for influencing smoking cessation because it was more personalized (it learned and adapted to a person's behavior) and could incorporate a considerably greater number of variables; however, larger trials would be needed to demonstrate the system's effectiveness.

Zeevi et al [28] used gradient boosting regression to predict the postmeal glycemic response of individuals in real life, according to blood parameters, dietary habits, anthropometrics, physical activity, and gut microbiota. The results from this randomized controlled study with 24 participants showed that a personalized diet based on postmeal glycemic predictions could statistically significantly modify elevated postprandial blood glucose (*P*<.05 for predicting low levels of blood glucose ["good diet"] vs high levels of blood glucose ["bad diet"], which was comparable with diets selected by experts). The authors reported that their approach could be used in nutritional interventions for controlling or preventing disorders associated with poor glycemic control, such as obesity, diabetes, and nonalcoholic fatty liver disease. However, evaluation periods of months or even years would be needed first to clearly indicate the effectiveness of the proposed algorithm.

Discussion

Principal Findings

This review is, to our knowledge, the first to systematically examine the features and outcomes of digital health interventions incorporating machine learning that were implemented and assessed in real-life studies [17]. With this aim in mind, we differentiated our review from previous investigations that focused only on the broader use of artificial intelligence in medicine in the context of specific diseases [29,30], machine learning techniques [31,32], or risk prediction models, such as through mining of electronic health records [33,34], and did not consider real-life evaluation of the respective interventions. The need to demonstrate evidence of an intervention's effectiveness in the real world has been highlighted in several other studies [35-37]. Our main finding is that most of the digital health interventions showed significantly positive health outcomes for patients or healthy individuals, which demonstrates the virtue of machine learning applications in actual clinical practice. However, given the small number of studies identified in this review and their considerable limitations highlighted above, further work is warranted to demonstrate the effectiveness of digital interventions relying on machine learning applications in real-life medical care.

Our review found 8 different cases of machine learning applications in a real-life setting: depression prediction and management, speech recognition for people with speech disabilities, self-efficacy for weight loss, detection of changes in biopsychosocial condition of patients with multiple morbidity, stress management, treatment of phantom limb pain, smoking cessation, and personalized nutrition based on glycemic response. The reviewed studies had several implications for clinical practice, such as better engagement of patients with interventions [22], the identification of risk for hospitalization [25], or the introduction of novel treatment methods [27]. Among the studies, those for speech recognition of people with speech disabilities [24] and notification delivery for stress management [26] clearly reported insignificant outcomes, whereas 6 studies showed significant outcomes, but they were of low to moderate methodological quality. Only 3 studies were in the form of a randomized controlled trial, which limited the ability to fully identify the added value of machine learning-enabled interventions compared with standard care. To this end, further rigorous studies with adequately powered samples (recruiting considerably more participants than the average number of 71 participants found in this review) are needed, which would generate the evidence base for the effectiveness of machine learning in clinical practice. To that effect, large trials and publicly accessible databases that have become available over the last few years, such as the UK BioBank and the Physionet database, are providing rich resources that could facilitate insights.

The delivery of motivational messages [21,26] or stories [22] for health behavior change and engagement seems to be an emerging area of digital health interventions incorporating machine learning. These studies also demonstrated the latest efforts to promote individuals' personalized self-management and to put them at the center of health care [38]. Considering the effectiveness of tailored messaging in influencing health behavior change [39], further research in this area is warranted.

The surprisingly small number of identified pragmatic studies in our review might raise some concerns and indicates the substantial challenge of systematically evaluating digital health interventions that incorporate machine learning [40]. In this context, the retrospective validation of algorithms and models, given the availability of one or more datasets, constitutes only the first step in the evaluation process [28]. The second step involves the integration of the algorithms and models within a digital health tool, such as mobile phone–based tools [23], internet-based tools [14], or an aid device [24]. The third step requires the assessment of the developed tool as a digital health intervention in a real-life research setting (eg, through a randomized controlled trial), together with patients or health professionals, or both [28,41]. The final step would be the monitoring of actual uptake and use of the intervention in real-world settings and outside of a research setting [42], which is, however, rarely reported [43]. Admittedly, this process is challenging and anything but trivial. It requires a significant amount of time and resources, which might not always be available, and multidisciplinary collaboration among experts in different fields, such as engineering, computer science, behavioral science, and medicine, which might not be straightforward. However, such synergistic collaborative approaches are likely necessary in the development of evidence-based, sustainable, and impactful digital health interventions [44,45].

Limitations

We used the term *machine learning*, along with broader terms such as *data mining* and *artificial intelligence*, for our literature search, rather than keywords for specific machine learning algorithms or domains relevant to digital health, such as *telemedicine*. This might have inadvertently omitted studies that could have contributed to the progress made in machine learning applications for digital health. We combined the aforementioned terms with the generic term *health*, aiming to conduct a broad search within the provided boundaries and to include the most pertinent articles relevant to digital health. We searched for articles in a limited number of databases (ie, PubMed and Scopus), which nevertheless are two of the most widely used databases internationally [46]. We did not hand search any studies reported in other reviews or the included studies, and we did not assess the interrater reliability. A meta-analysis was not possible due to the heterogeneity of the included studies.

Conclusion

Our review showed that real-life digital health interventions incorporating machine learning can be useful and effective. Considering the small number of studies examined in this review and their limitations, further evidence of the clinical usefulness of machine learning in health service delivery is needed. We encourage researchers to move beyond the retrospective validation of machine learning models, by integrating their models within appropriately designed digital health tools and evaluating their tools in rigorous studies conducted in real-life settings.

Acknowledgments

The study was supported by the Wellcome Trust through Centre Grant No. 098461/Z/12/Z from the University of Oxford Sleep & Circadian Neuroscience Institute (SCNi). It was also supported by Health Data Research UK, which receives funding from HDR UK Ltd (HDR-5012) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), the British Heart Foundation, and the Wellcome Trust. The funders had no role in the study or the decision to submit this work to be considered for publication.

Authors' Contributions

AKT was responsible for conducting the study; AKT and AT reviewed the literature and assessed the quality of the included studies; AKT synthesized the literature according to the described methodology; and AKT wrote the first draft of the manuscript and AT contributed to the final version. Both authors read and agreed to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA checklist.

[PDF File (Adobe PDF File), 500KB - jmir_v21i4e12286_app1.pdf]

Multimedia Appendix 2

Characteristics of included studies and implications for clinical practice.

[PDF File (Adobe PDF File), 40KB - jmir_v21i4e12286_app2.pdf]

References

1. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
2. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: The SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
3. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
4. 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]
5. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
6. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
7. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-116 [FREE Full text] [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]
8. 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. [doi: 10.1016/j.jpainsymman.2018.02.016] [Medline: 29496537]
9. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12(4):e0174944 [FREE Full text] [doi: 10.1371/journal.pone.0174944] [Medline: 28376093]
10. Kelly J, Gooding P, Pratt D, Ainsworth J, Welford M, Tarrier N. Intelligent real-time therapy: harnessing the power of machine learning to optimise the delivery of momentary cognitive-behavioural interventions. J Ment Health 2012 Aug;21(4):404-414. [doi: 10.3109/09638237.2011.638001] [Medline: 22251028]
11. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2007 Feb 11;2:59-77 [FREE Full text] [Medline: 19458758]
12. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, Alzheimer's Disease Neuroimaging Initiative. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front Neurosci 2015;9:307 [FREE Full text] [doi: 10.3389/fnins.2015.00307] [Medline: 26388719]
13. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]

14. Triantafyllidis A, Filos D, Buys R, Claes J, Cornelissen V, Kouidi E, et al. Computerized decision support for beneficial home-based exercise rehabilitation in patients with cardiovascular disease. Comput Methods Programs Biomed 2018 Aug;162:1-10. [doi: 10.1016/j.cmpb.2018.04.030] [Medline: 29903475]
15. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
16. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
17. Agboola S, Hale TM, Masters C, Kvedar J, Jethwani K. “Real-world” practical evaluation strategies: a review of telehealth evaluation. JMIR Res Protoc 2014;3(4):e75 [FREE Full text] [doi: 10.2196/resprot.3459] [Medline: 25524892]
18. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
20. Mohammadi E, Thelwall M, Haustein S, Larivière V. Who reads research articles? An altmetrics analysis of Mendeley user categories. J Assoc Inf Sci Technol 2015 Apr 29;66(9):1832-1846. [doi: 10.1002/asi.23286]
21. Sadasivam RS, Borglund EM, Adams R, Marlin BM, Houston TK. Impact of a collective intelligence tailored messaging system on smoking cessation: the Perspect randomized experiment. J Med Internet Res 2016 Nov 08;18(11):e285 [FREE Full text] [doi: 10.2196/jmir.6465] [Medline: 27826134]
22. Manuvinakurike R, Velicer WF, Bickmore TW. Automated indexing of Internet stories for health behavior change: weight loss attitude pilot study. J Med Internet Res 2014;16(12):e285 [FREE Full text] [doi: 10.2196/jmir.3702] [Medline: 25491243]
23. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
24. Hawley MS, Cunningham SP, Green PD, Enderby P, Palmer R, Sehgal S, et al. A voice-input voice-output communication aid for people with severe speech impairment. IEEE Trans Neural Syst Rehabil Eng 2013 Jan;21(1):23-31. [doi: 10.1109/TNSRE.2012.2209678] [Medline: 22875259]
25. Martin CM, Vogel C, Grady D, Zarabzadeh A, Hederman L, Kellett J, et al. Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR). J Eval Clin Pract 2012 Dec;18(6):1226-1234. [doi: 10.1111/j.1365-2753.2012.01880.x] [Medline: 22816797]
26. Morrison LG, Hargood C, Pejovic V, Geraghty AWA, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
27. Ortiz-Catalan M, Guðmundsdóttir RA, Kristoffersen MB, Zepeda-Echavarria A, Caine-Winterberger K, Kulbacka-Ortiz K, et al. Phantom motor execution facilitated by machine learning and augmented reality as treatment for phantom limb pain: a single group, clinical trial in patients with chronic intractable phantom limb pain. Lancet 2016 Dec 10;388(10062):2885-2894. [doi: 10.1016/S0140-6736(16)31598-7] [Medline: 27916234]
28. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015 Nov 19;163(5):1079-1094 [FREE Full text] [doi: 10.1016/j.cell.2015.11.001] [Medline: 26590418]
29. Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data mining algorithms and techniques in mental health: a systematic review. J Med Syst 2018 Jul 21;42(9):161. [doi: 10.1007/s10916-018-1018-2] [Medline: 30030644]
30. 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]
31. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 2012 Apr;36(4):1140-1152. [doi: 10.1016/j.neubiorev.2012.01.004] [Medline: 22305994]
32. 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]
33. Goldstein BA, Pencina MJ, Montez-Rath ME, Winkelmayer WC. Predicting mortality over different time horizons: which data elements are needed? J Am Med Inform Assoc 2017 Dec;24(1):176-181 [FREE Full text] [doi: 10.1093/jamia/ocw057] [Medline: 27357832]
34. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]

35. Dunkley AJ, Bodicoat DH, Greaves CJ, Russell C, Yates T, Davies MJ, et al. Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. Diabetes Care 2014 Apr;37(4):922-933. [doi: 10.2337/dc13-2195] [Medline: 24652723]
36. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
37. Aziz Z, Absetz P, Oldroyd J, Pronk NP, Oldenburg B. A systematic review of real-world diabetes prevention programs: learnings from the last 15 years. Implement Sci 2015 Dec 15;10:172 [FREE Full text] [doi: 10.1186/s13012-015-0354-6] [Medline: 26670418]
38. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [Medline: 12867348]
39. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
40. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: WHO; 2016.
41. Claes J, Buys R, Woods C, Briggs A, Geue C, Aitken M, et al. PATHway I: design and rationale for the investigation of the feasibility, clinical effectiveness and cost-effectiveness of a technology-enabled cardiac rehabilitation platform. BMJ Open 2017 Jun 30;7(6):e016781 [FREE Full text] [doi: 10.1136/bmjopen-2017-016781] [Medline: 28667228]
42. Michelson D, Davenport C, Dretzke J, Barlow J, Day C. Do evidence-based interventions work when tested in the “real world?” A systematic review and meta-analysis of parent management training for the treatment of child disruptive behavior. Clin Child Fam Psychol Rev 2013 Mar;16(1):18-34. [doi: 10.1007/s10567-013-0128-0] [Medline: 23420407]
43. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
44. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
45. Cao L, Ramesh B. Agile requirements engineering practices: an empirical study. IEEE Softw 2008 Jan;25(1):60-67. [doi: 10.1109/MS.2008.1]
46. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]

Abbreviations

EPHPP: Effective Public Health Practice Project
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

---

Edited by CL Parra-Calderón; submitted 21.09.18; peer-reviewed by K Cato, J Claes, M Larsen; comments to author 12.11.18; revised version received 07.01.19; accepted 26.01.19; published 05.04.19.

Please cite as:
Triantafyllidis AK, Tsanas A
Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature
J Med Internet Res 2019;21(4):e12286
URL: https://www.jmir.org/2019/4/e12286/
doi:10.2196/12286
PMID:30950797

Original Paper

Detecting Developmental Delay and Autism Through Machine Learning Models Using Home Videos of Bangladeshi Children: Development and Validation Study

Qandeel Tariq[1,2*], MA; Scott Lanyon Fleming[2*], BS; Jessey Nicole Schwartz[1,2], BA; Kaitlyn Dunlap[1,2], BSc, MRES; Conor Corbin[2], BS; Peter Washington[2], MS; Haik Kalantarian[1,2], PhD; Naila Z Khan[3], PhD, MBBS, FCPS; Gary L Darmstadt[1,4], PhD; Dennis Paul Wall[1,2], PhD

[1]Division of Systems Medicine, Department of Pediatrics, Stanford University, Palo Alto, CA, United States
[2]Department of Biomedical Data Science, Stanford University, Palo Alto, CA, United States
[3]Dhaka Shishu Children's Hospital, Dhaka, Bangladesh
[4]Division of Neonatal and Developmental Medicine, Department of Pediatrics, Stanford University, Palo Alto, CA, United States
*these authors contributed equally

Corresponding Author:
Dennis Paul Wall, PhD
Division of Systems Medicine
Department of Pediatrics
Stanford University
1265 Welch Road
Palo Alto, CA, 94305
United States
Phone: 1 6173946031
Email: dpwall@stanford.edu

Abstract

Background: Autism spectrum disorder (ASD) is currently diagnosed using qualitative methods that measure between 20-100 behaviors, can span multiple appointments with trained clinicians, and take several hours to complete. In our previous work, we demonstrated the efficacy of machine learning classifiers to accelerate the process by collecting home videos of US-based children, identifying a reduced subset of behavioral features that are scored by untrained raters using a machine learning classifier to determine children's "risk scores" for autism. We achieved an accuracy of 92% (95% CI 88%-97%) on US videos using a classifier built on five features.

Objective: Using videos of Bangladeshi children collected from Dhaka Shishu Children's Hospital, we aim to scale our pipeline to another culture and other developmental delays, including speech and language conditions.

Methods: Although our previously published and validated pipeline and set of classifiers perform reasonably well on Bangladeshi videos (75% accuracy, 95% CI 71%-78%), this work improves on that accuracy through the development and application of a powerful new technique for adaptive aggregation of crowdsourced labels. We enhance both the utility and performance of our model by building two classification layers: The first layer distinguishes between typical and atypical behavior, and the second layer distinguishes between ASD and non-ASD. In each of the layers, we use a unique rater weighting scheme to aggregate classification scores from different raters based on their expertise. We also determine Shapley values for the most important features in the classifier to understand how the classifiers' process aligns with clinical intuition.

Results: Using these techniques, we achieved an accuracy (area under the curve [AUC]) of 76% (SD 3%) and sensitivity of 76% (SD 4%) for identifying atypical children from among developmentally delayed children, and an accuracy (AUC) of 85% (SD 5%) and sensitivity of 76% (SD 6%) for identifying children with ASD from those predicted to have other developmental delays.

Conclusions: These results show promise for using a mobile video-based and machine learning–directed approach for early and remote detection of autism in Bangladeshi children. This strategy could provide important resources for developmental health in developing countries with few clinical resources for diagnosis, helping children get access to care at an early age. Future research aimed at extending the application of this approach to identify a range of other conditions and determine the population-level burden of developmental disabilities and impairments will be of high value.

(J Med Internet Res 2019;21(4):e13822) doi:10.2196/13822

KEYWORDS

autism; autism spectrum disorder; machine learning; developmental delays; clinical resources; Bangladesh; Biomedical Data Science

Introduction

Autism spectrum disorder (ASD) is a heterogeneous developmental disorder that includes deficits in social communication, repetitive behaviors, and restrictive interests, all of which lead to significant social and occupational impairments throughout the lifespan. Autism is one of the fastest growing developmental disorders in the United States [1], affecting 1 in 59 children [2]. Although the global autism prevalence is largely unknown, the prevalence is estimated to be between 0.15% and 0.8% among children in developing countries such as Bangladesh, with a higher prevalence in urban centers (eg, 3% in Dhaka) [3]. These numbers only represent a fraction of the actual cases, as most cases in semiurban and rural areas go unnoticed due to a dearth of resources. The disparity between urban and rural prevalence may reflect poorly understood risk factors or clinical resources in high-income areas along with higher awareness among urban parents about developmental delays [4]. More accessible and wide-scale screening is needed to accurately estimate ASD prevalence in remote parts of Bangladesh and other countries.

The current models for diagnosing autism in Bangladesh, as in the United States, are often administered by trained clinical professionals using standard assessments [5]. Empirically validated diagnostic tools like the Autism Diagnostic Observation Schedule (ADOS) [6] and Autism Diagnostic Interview (ADI-R) [7] are not always used in different countries, particularly in developing countries, as these tools are expensive, require trained clinicians to administer, and may be limited by available translations and cultural adaptations [4]. For countries with limited ASD resources like Bangladesh, obtaining a diagnosis, which is essential for receiving an intervention and improving outcomes, is difficult. There is a pressing need to further develop open-source tools that do not require extensive training and professional certification and have high cross-cultural validity for autism screening globally [4]. Previous work has shown the feasibility and efficacy of assessing developmental delay using rapid assessment tools delivered by professionals with limited clinical expertise in the home [5]. There is potential to extend the reach of assessment tools and decrease health care disparity, especially in developing and rural countries, by using machine learning and mobile technologies.

In our previous works, we have developed tools for rapid mobile detection of ASD in short home videos of US children by using supervised machine learning approaches to identify minimal sets of behaviors that align with clinical diagnoses of ASD [8-15]. Features extracted in our minimally viable classifiers are accurately labeled by nonexpert raters (ie, noncertified clinical practitioners) in a short period of time (eg, <6 minutes). These labeled features can then be fed into our machine learning classifiers to determine the child’s autism risk. Tariq et al [14] used a dataset consisting of 162 videos (116 ASD, 46 neurotypical development [TD]) of US children to validate these classifiers. The top-performing classifier exhibited an accuracy of 92% (95% CI 88%-97%).

Additionally, an independent validation set consisting of 66 videos (33 ASD, 33 TD) was labeled by a separate set of video raters in order to validate the results. The top-performing classifier maintained similar results, achieving an overall accuracy of 89% (95% CI 81%-95%).

The current study aimed to show generalizability of video-based machine learning procedures for ASD detection that have established validity among US-based children [14] in Bangladesh. Specifically, our study aimed to determine the performance and accuracy of this same video machine learning procedures on videos of Bangladeshi children under the age of 4 years. This sample was drawn from a population diagnosed with ASD and another population with other speech and language conditions (SLCs), but not ASD. Additionally, we compared the features that are most important for accurate classification of children from Bangladesh and created several machine learning models that can be generalized to different cultures.

Methods

Data Collection

The study received ethical clearance under Dr Naila Khan from the Bangladesh Institute of Child Health, Dhaka Shishu Children’s Hospital (DSH) and the Stanford University Institutional Review Board. We aimed to recruit 150 children for this study: 50 with ASD, 50 with an SLC, and 50 with neurotypical development (TD). All participants were recruited after they provided consent (in Bengali language) for participation at the DSH, and their children were screened for the presence of ASD or SLC. Participants were enrolled if they were parents above 18 years of age, had a child between the ages of 18 months and 4 years, could attend an appointment at the DSH to complete the study procedures, and were willing to submit a brief video of their child to the study team. Enrolled families provided demographic information (see Table 1 in the Results section).

Brief videos (2-5 minutes) were recorded during evaluation of the children who presented to the Child Development Center of the Bangladesh Institute of Child Health with neurodevelopmental concerns. We administered the Modified Checklist for Autism in Toddlers (Bangla version [16]) to all children to identify the presence of ASD, and all children underwent additional clinical evaluations by a developmental psychologist and a child health physician in order diagnose ASD, SLC, or TD, as described previously [5]. We also administered the ADOS for 28 of the 50 children identified with ASD; ADOS could not be completed in the remaining 22 children diagnosed with ASD because their families were unable to commit to the time required to complete the assessment, a common problem for families in low-resource areas [4].

Acquired videos and supporting demographic measures were securely sent from DSH to Stanford University. Videos were assessed for quality by trained clinical researchers at Stanford University. Criteria included video, sound, and image quality in addition to video length and content (ie, ensuring that the video was long enough to answer necessary questions, that the child was present in the video, etc). Furthermore, videos were assessed to meet the following criteria: (1) it captured the child's face and hands, (2) it involved social interaction or attempts of social interaction, and (3) it involved an interaction between the child and a toy/object.

Video Raters

Nine non-Bengali speaking US-based raters with no clinical training used a secure, HIPAA (Health Insurance Portability and Accountability Act)-compliant online website to watch the videos and answer a set of 31 multiple-choice questions corresponding to the behavioral features of autism [14]. Each rater completed a 1-hour training session with a senior analyst before scoring the videos. Senior analysts conducted rater quality checks by comparing a subset of 10 video scores to "gold standard" scores. These "gold standard" scores were agreed upon by two clinical research coordinators who each had several years of experience with children with autism.

Source Classifiers Trained on Clinical Data for Reduce-to-Practice Testing

We assembled eight published machine learning classifiers to test their viability for use in the rapid mobile detection of autism through the use of short home videos of US children [14]. For all eight models, the source of training and validation data was item-level medical records of US children, which contained either the ADOS or ADI-R outcome data on all participants. The ADOS has several modules containing approximately 30 features that correspond to the developmental level of the individual under assessment. These features are assessed based on how a child interacts with a clinical practitioner administering the exam. The ADI-R is a parent-directed interview that includes >90 elements asked of the parent, with multiple choices for answers. Each model was trained on item-level outcomes from the administration of either the ADOS or ADI-R and optimized for accuracy, sparsity of features, and interpretability in previous publications [8-15]. All these classifiers have been validated with US home videos (total: n=162, ASD: n=116, non-ASD: n=46) [14]. The top three performing classifiers in this dataset were chosen for validation of the videos collected from DSH in Bangladesh to test the accuracies of these models across cultures.

Stacked Classifiers With Rater-Adaptive Weighting

In an effort to improve the results on the Bangladeshi dataset after attempting to validate previously built classifiers on these data, we constructed new classifiers while controlling for potential noise resulting from inaccurate ratings and constructed separate layers for each step of the classification for a streamlined approach. Our dataset contained three classes—TD, ASD, and SLC—assigned by screening via clinical evaluation at the DSH [5]. By implementing a layered approach to classification—first distinguishing general developmental delays (including ASD and SLC) from TD and then distinguishing ASD from SLCs—we were able to broaden the detection capabilities to more generally classify the presence of other developmental delays in addition to ASD specifically.

Rater Weighting

Given the raters' lack of formal clinical training, we hypothesized that some raters might be more adept at identifying certain risk factors in some videos than others. Regardless of whether these interrater differences in identification accuracy for certain subsets of behaviors arise naturally or by chance, we hypothesized that this heterogeneous rater performance could be leveraged to yield increased model performance. For example, if one rater is especially capable of labeling a child's level of eye contact and another rater does a poor job of rating eye contact but excels at rating language ability, then a model trained on each individual rater's labels alone might perform poorly; however, an ensemble that considers the outputs of both rater's models could perform substantially better. Achieving this improved performance is the focus of our proposed novel rater-adaptive weighting scheme.

For each of the three raters in the dataset, we trained a Random Forest classifier to predict a child's class label (TD, SLC, or ASD) based on the rater's annotations of that child's behavior in a given video. The Random Forest classifier adapts to each rater's expertise and labeling patterns; a basic analysis revealed that each rater had a different feature set that they rated well. In addition to (and, in part, because of) interrater differences in the labeling ability, each rater's model had varying levels of accuracy. We wanted the ensemble to weigh the predictions from the most accurate rater models more heavily. Therefore, we first trained and calculated the accuracy of each rater's model relative to a majority vote baseline and then used that difference to up- or downweigh that rater's vote relative to the other raters' votes.

Specifically, we let z_j represent the difference in accuracy of rater *j* 's model relative to the majority vote baseline. Then, after calculating z_j for each rater j=1, 2...K, we pass these values into the softmax function to generate rater-specific weights:

This ensures that all the raters' weights collectively sum up to 1, so that the ensemble prediction will be a linear combination of each rater's predictions. Using these weights, the final ensemble prediction for child *i* is calculated by multiplying each rater model's predicted probability for the target class (eg, atypical development or ASD) by the corresponding rater-specific weight and adding the weighted raters' predicted probabilities together. More specific details can be found in Multimedia Appendix 1.

Stacking Classifiers to Distinguish Between Typical/Atypical Development and Autism Spectrum Disorder/Speech and Language Conditions

In order to reflect the differences in both the conceptualization and use cases of predicting (1) TD vs atypical development and (2) ASD from other developmental delays, we decided to create a stacked approach to classification. In the first layer, we built classifiers to distinguish between TD and atypical development (ASD/other SLCs). The cases classified as atypical from the first layer were then used as input for the second layer to distinguish between ASD and other SLCs.

We wanted to optimize the model for sensitivity in the first layer to ensure no atypical case was misclassified. In the second layer, we wanted to optimize for both sensitivity and specificity, so that children with ASD would be effectively distinguished from children with other development delays. After training these classifiers for each rater, we tested them on the held-out test set and aggregated rater scores using the rater weights calculated in the previous step. For each of these layers, we used a three-fold cross-validation approach to select the training and test sets randomly in order to ensure that the accuracy reported is stable across different splits.

Feature Importance

To determine the impact of each video's annotations on the classifier's predicted label for that video, we used a recently developed method for efficiently calculating approximate Shapley values [17]. Shapley values are traditionally used in coalitional game theory to determine how to optimally distribute gains earned from cooperative effort. The same idea can be extended to machine learning in order to rank features for nonlinear models such as Random Forests. In the machine learning adaptation of Shapley values, feature values "cooperate" to impact a machine learning model's output, which in this case is the predicted probability of a child's video being classified as TD, ASD, or SLC. For each video, Shapley values capture both the magnitude of importance for each feature value as well as the direction in which the feature value "pushes" the final predicted class probability. More precisely, if we let $\Phi_k(F_j^*, x^{(i)})$ be the impact (Shapley value) of the *k* th feature for video *i* with feature vector $x^{(i)}$ on the output of model F_j^*, then the Shapley value formulation guarantees that

In other words, any video's final predicted class probability is the average predicted class probability of the dataset plus all the Shapley values associated with each element of that video's input vector. This property, called local accuracy, indicates that the feature importance can be easily measured and compared. Additionally, because each video, feature, and model triple is associated with a single scalar-valued feature importance, we can understand how each annotation for each child's video affected his/her predicted probability of TD/ASD/SLC at an individual level and estimate a feature's overall importance to the model by summing up the absolute values of that feature's Shapley values over all videos. The features with the highest sum of absolute Shapley values are considered the most important to the model. Finally, given the way in which we ensembled individual raters' models, we can extract Shapley values for the multirater ensemble by employing the same weights. Specifically, we can employ the following equation:

To test whether our classifier's decisions align with clinical intuition, we calculated Shapley values for the 159 videos for the second layer of the classifier when distinguishing ASD from non-ASD.

Comparing Bangladeshi and US Results

In order to determine the generalizability of one dataset's characteristics to the other, we trained logistic regression classifiers with elastic net regularization for the Bangladeshi data and US data to predict ASD from the non-autism class. We trained the model on the Bangladeshi data and tested the model on the US data and vice versa. For both classifiers, we randomly split the dataset into training and testing, reserving 20% for the latter while using cross-validation on the training set to tune hyperparameters associated with elastic net regularization. Note that while traditional logistic regression seeks to find a set of model coefficients, β, that minimizes the logarithmic loss (we will denote this loss as where represents the model's predictions when the model is parameterized by β), logistic regression with elastic net regularization seeks to minimize the logarithmic loss plus a regularization term:

Here, the first sum corresponds to an L2-loss, the second sum corresponds to an L1-loss, ρ is a hyperparameter governing the balance between the two losses, and α is the second hyperparameter determining the overall strength of regularization. Incorporating this regularization into the logistic regression loss yields several benefits, including more parsimonious and interpretable models and better predictive performance, especially when two or more of the predictor variables are correlated [18]. We used cross-validation for model hyperparameter tuning by performing a grid search with different values of α (varying penalty weights) and ρ (the mixing parameter determining how much weight to apply to L1 versus L2 penalties) [19,20-21]. Based on the resulting area under the curve (AUC) and accuracy from each combination, we selected the top-performing pair of hyperparameters. Using this pair, we trained the model using logistic regression and balanced class weights to adjust weights that were inversely proportional to class frequencies in the input data, which helps account for class imbalance. After determining the top-ranked features based on the trained model and the resulting coefficients, we validated the model on the reserved test set. The behavioral features that were selected most often during the hyperparameter tuning phase across different cross-folds were compared between US and Bangladeshi models to determine which features have a greater significance and whether they align between the two models.

Software

Analyses were performed in Python 3.6.7; we used pandas 0.23.4 to prepare the data for analysis [20]. The classification models described were trained and evaluated using the scikit-learn 0.20.0 package [21]. Hyperparameters for each rater model were tuned using the hyperopt 0.1.1 package [22]. Shapley value estimates were calculated using the shap 0.24.0 package [23]. Plots were generated using matplotlib 3.0.1 [24].

Results

Data Collection

We collected 159 videos in total: 55 videos were of children with ASD, 50 were of children with SLC, and 54 were of children with TD. The parent-submitted home videos were an average of 3 minutes 11 seconds long (SD 1 minute 57 seconds). Of the 159 videos submitted, all were manually inspected and found to be of good, scorable quality in terms of length, resolution, and content. Demographic data were missing for 9 subjects, who were excluded from analysis; all other data were complete. Video rating staff were able to rate all videos. Table 1 outlines the diagnosis and demographic breakdown for 150 of the 159 videos included in the dataset.

Results of Source Classifiers Trained on Clinical Data for Reduce-to-Practice Testing

We first sought to distinguish AD from non-ASD cases. Our top performing classifiers from our previous analysis of the videos from 162 US children [14] were validated on the Bangladeshi dataset. We tested across different train-test splits and achieved a maximum AUC of 0.75 (SD 0.06; Figure 1). In order to improve classifier performance, we next shifted to the development of stacked classifiers.

Results From Stacked Classifiers With Rater-Adaptive Weightings

Since we used a three-fold cross-validation approach, we trained and tested the models for each of the raters across three different splits. The training set consisted of 114 randomly selected videos, and the average demographic information for the three splits for the training set was as follows: average age, 2 years 7 months (SD 7 months); proportion of males, 64%; proportion of children with TD, 34%; proportion of children with SLC, 31%; and proportion of children with ASD, 35%. The demographic information for the test set for layer 1 (distinguishing TD from ASD/SLC) and layer 2 (distinguishing ASD from SLC) can be found in Table 2.

Table 1. Participant demographics collected from Dhaka Shishu Hospital, Bangladesh.

Demographic	Full cohort (N=150)	ASD[a] cohort (N=50)	TD[b] cohort (N=50)	SLC[c] cohort (N=50)
Age (years), mean (SD)	2.55 (0.62)	2.51 (0.70)	2.40 (0.59)	2.73 (0.51)
Gender (male), n (%)	90 (60)	36 (72)	23 (62)	31 (46)
Preterm (ie, <37 weeks), n (%)	11 (0.7)	5 (10)	0 (0)	6 (12)
Family income in taka[d], n (%)				
1,000-10,000	16 (10.7)	0 (0)	16 (32)	0 (0)
>10,000-30,000	33 (22)	2 (4)	21 (42)	10 (20)
>30,000	101 (67.3)	48 (96)	13 (26)	40 (80)
Residence, n (%)				
Urban	139 (92.7)	50 (100)	50 (100)	39 (78)
Semiurban	8 (5.3)	0 (0)	0 (0)	8 (16)
Rural	3 (2)	0 (0)	0 (0)	3 (6)
Religion, n (%)				
Muslim	141 (94)	44 (88)	49 (98)	48 (96)
Hindu	6 (4)	4 (8)	0 (0)	2 (4)
Christian	1 (0.01)	1 (2)	0 (0)	0 (0)
Buddhist	2 (0.01)	1 (2)	1 (2)	0 (0)
Stunted growth, n (%)				
Missing stunting information	60 (40)	4 (8)	50 (100)	6 (12)
No stunting	49 (32.7)	30 (60)	0 (0)	19 (48)
Stunting	41 (27.3)	16 (32)	0 (0)	25 (50)
Clinical evaluations, mean (SD)				
MCHAT[e] total score		**13.5 (3.04)**	**2 (0)**	**0.08 (0.57)**
ADOS[f,g] score				
Social affect	N/A[h]	11.57 (5.30)	N/A	N/A
Restricted and repetitive behavior	N/A	3.46 (3.29)	N/A	N/A
Composite	N/A	5.14 (2.08)	N/A	N/A
SLC diagnosis				
Receptive language delay	N/A	N/A	N/A	2 (4)
Expressive language delay	N/A	N/A	N/A	5 (10)
Both receptive and expressive language delay	N/A	N/A	N/A	37 (74)
Receptive and expressive language disorder	N/A	N/A	N/A	6 (12)

[a] ASD: autism spectrum disorder.

[b] TD: neurotypical development.

[c] SLC: speech and language condition.

[d] 1 US $=84 taka.

[e] MCHAT: Modified Checklist for Autism in Toddlers

[f] ADOS: Autism Diagnostic Observation Schedule.

[g] ADOS was only performed on a subset of 28 children with ASD.

[h] N/A: not available.

Figure 1. Results from the top performing classifiers trained on US clinical score sheet data and tested on Bangladeshi data with an objective to distinguish between ASD and non-ASD. ROC: receiver operating characteristic; AUC: area under the curve; ASD: autism spectrum disorder.

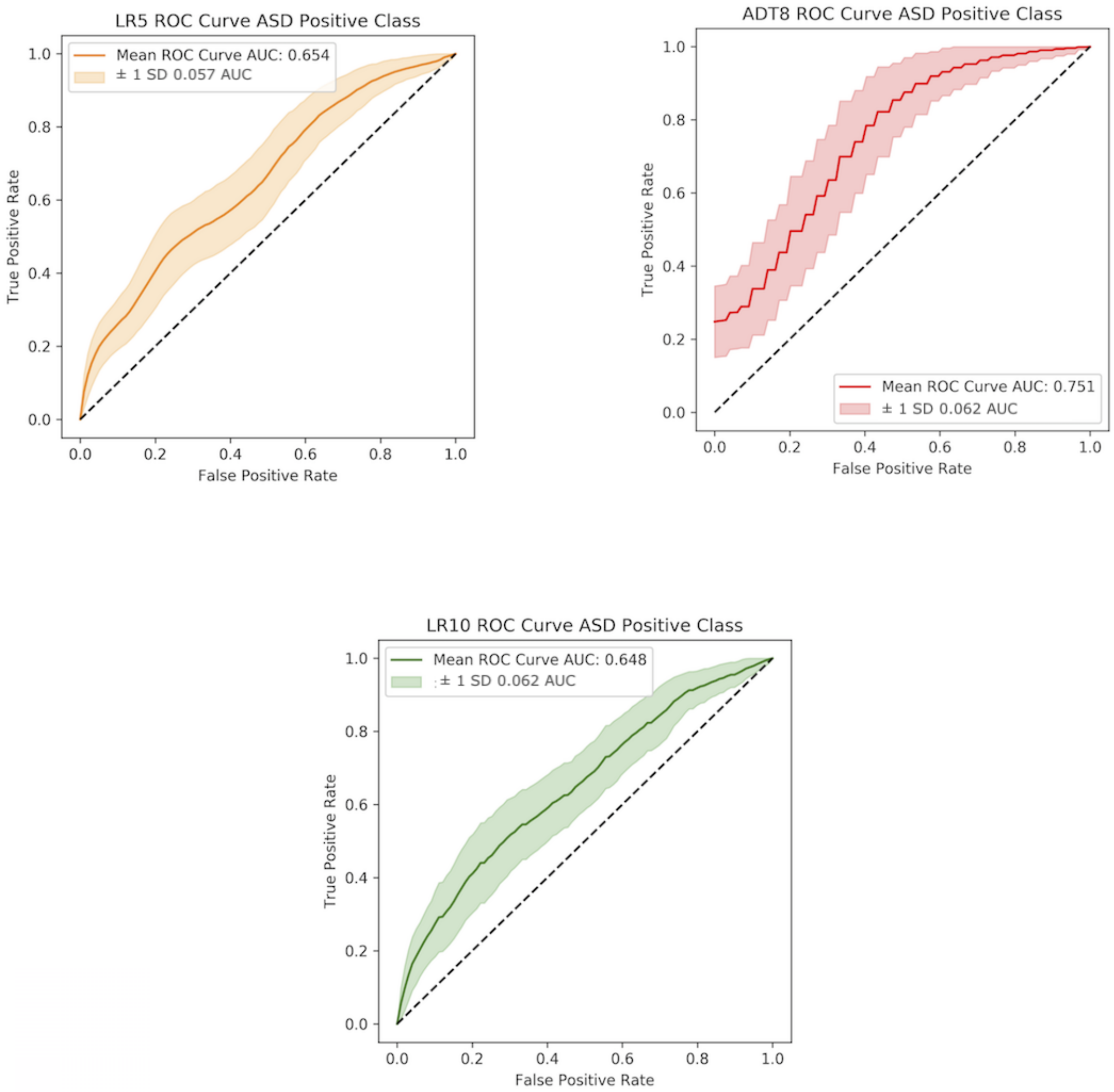


Table 2. Average demographic information of the test set calculated by testing the model on 45 videos for both layers.

Demographic	Layer 1 (distinguishing TD[a] from ASD[b]/SLC[c])	Layer 2 (distinguishing ASD from SLC)
Age (years), average (SD)	2 years 7 months (5 months)	2 years 6 months (3 months)
Proportion of males, mean %	62	70
Proportion of TD children, mean %	33	22
Proportion of children with ASD, mean %	33	44
Proportion of children with SLC, mean %	33	34

[a] TD: neurotypical development.

[b] ASD: autism spectrum disorder.

[c] SLC: speech and language condition.

Figure 2. (A) ROC curve for layer 1 (distinguishing between children with TD and children with ASD or SLC). (B) ROC curve for layer 2 (distinguishing between ASD and SLC). ASD: autism spectrum disorder; AUC: area under the curve; SLC: speech and language condition; TD: neurotypical development; ROC: receiver operating characteristic.

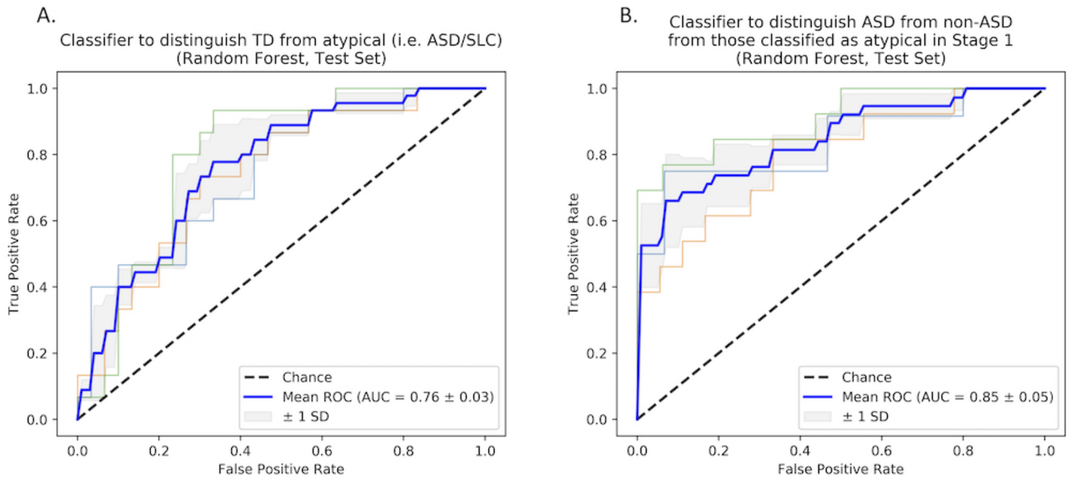


Table 3. Results from classifiers to distinguish among autism spectrum disorder, speech and language conditions, and neurotypical development. The results distinguish layer 1 (distinguishing neurotypical development from atypical conditions [autism spectrum disorder/speech and language conditions]) and layer 2 (distinguishing autism spectrum disorder from other delays [speech and language conditions]) from those classified as atypical in layer 1.

Classifier Layer	Sensitivity, % (SD)	Specificity, % (SD)	Unweighted average recall, % (SD)	Area under the curve, % (SD)	Accuracy, % (SD%)
Layer 1[a]	76 (SD 4)	58 (SD 3)	67 (SD 1)	76 (SD 3)	70 (SD 2)
Layer 2[b]	76 (SD 6)	77 (SD 24)	77 (SD 9)	85 (SD 5)	76 (SD 11)

[a] Distinguishing neurotypical development from autism spectrum disorder/speech and language conditions.

[b] Distinguishing autism spectrum disorder from other developmental delays (speech and language conditions).

Layer 1 of the stacked classifier, which sought to distinguish between children with TD from children with atypical development, achieved 76% (SD 4%) sensitivity and 58% (SD 3%) specificity with an AUC of 76% (SD 3%) and an accuracy of 70% (SD 2%; Figure 2 A). For layer 2, which distinguished ASD from other SLCs, the classifier performed with 76% (SD 6%) sensitivity, 77% (SD 24%) specificity with an AUC of 85% (SD 5%) and accuracy of 76% (SD 11%; Figure 2 B; Table 3).

Feature Importance

The most important features in our rater-adaptive ensemble for predicting ASD, as measured by the Shapley value, align with clinical intuition. Figure 3 shows the distribution of Shapley values across all participants for two of the features that were among the most important (as measured by mean absolute Shapley value) to our ensemble model's predictions. For example, for the feature corresponding to the child's level of eye contact, the value "rarely or never does this" contributes strongly to a classification of ASD and "exhibits clear, flexible gaze that is meshed with other communication" contributes the most to a non-ASD classification. Another feature that aligns with clinical intuition measures the child's repetitive interests and stereotyped behaviors—the feature value "behaviors observed the entire time," contributes strongly to the positive class (ASD), whereas "not observed" contributes strongly to the negative class (non-ASD; Figure 3).

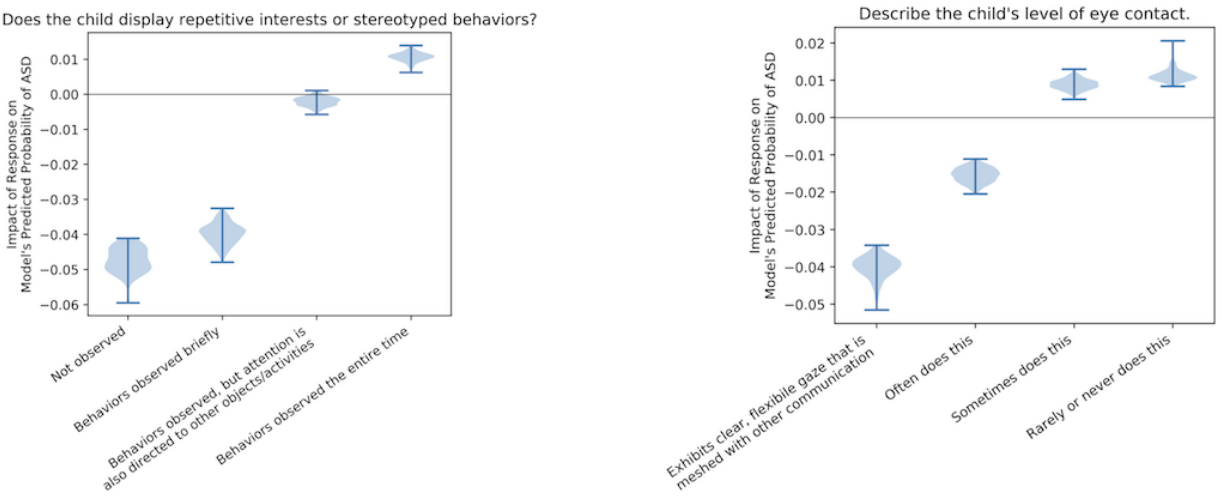


Figure 3. Shapley value distributions for two of the most important features in the rater-adaptive ensemble model. These features measure the child's stereotyped behaviors/repetitive interests and eye contact. They demonstrate that clinical intuition and the inner workings of our classifier align closely. ASD: autism spectrum disorder.

Comparison of Bangladeshi and US Results

For the classifier trained on the Bangladeshi data, the performance on the held-out test set (20% of Bangladeshi data) was 84.4% and its performance when validated on US data was 72.5% (Figure 4).

We trained a similar classifier on our dataset of 162 US videos and validated it on the Bangladeshi data (Figure 5). The classifier performed with a 94.2% accuracy when tested on the held-out test set from US videos. The classifier's accuracy dropped significantly when validated on the Bangladeshi data, reaching around 54%.

While performing hyperparameter tuning on these classifiers, we conducted further analysis to determine which of the behavioral features were selected most often for each cross-fold of US videos and Bangladeshi videos in order to draw a comparison. It is apparent from Figure 6 A and 6B that the features being selected are quite similar between the two datasets, with some minor differences. The features *understands language*, *sensory seeking*, *calls attention to objects*, and *stereotyped interests and actions* are highly ranked by models trained on either of the datasets. *Responsiveness*, *developmental delay*, *social participation*, and *stereotyped speech* are selected more often for US data and less so for Bangladeshi data. The opposite is true for *eye contact*.

Figure 4. Logistic regression (Elastic Net penalty) classifier, trained on Bangladeshi data and tested on US data as well as a held-out test set of the Bangladeshi data. AUC: area under the curve.

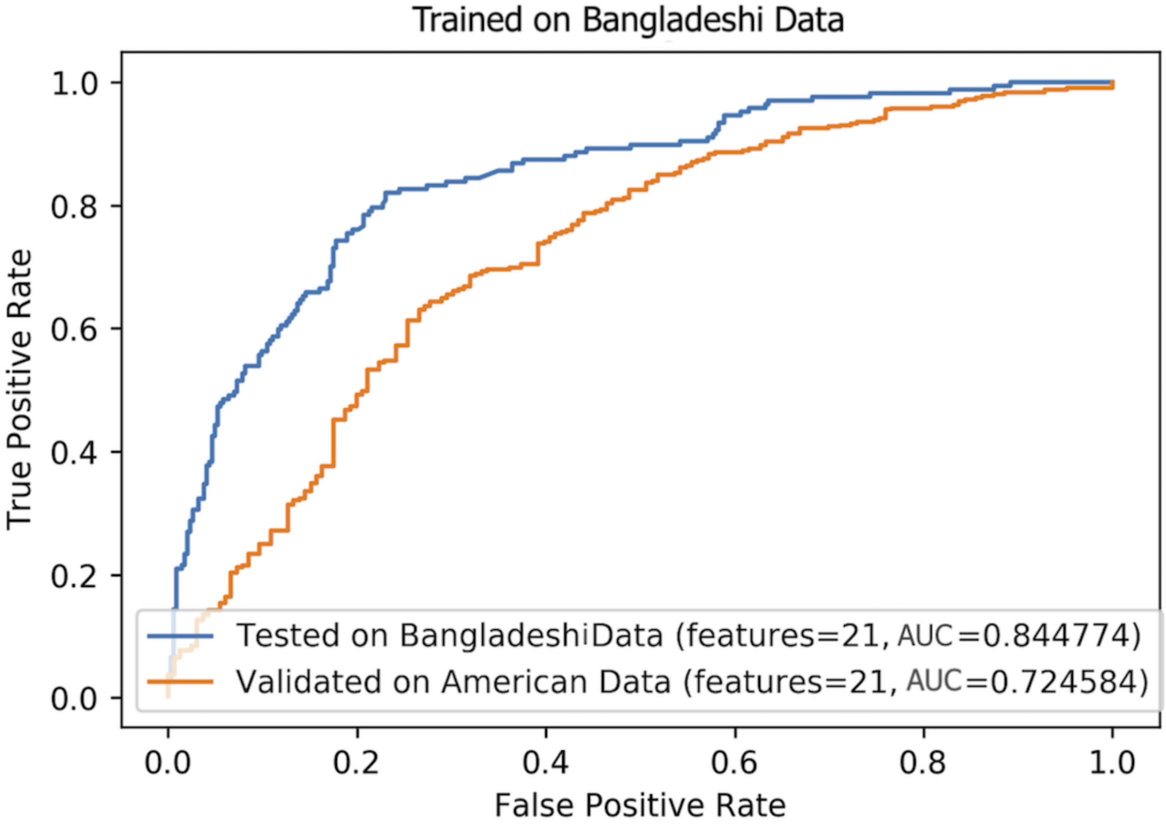


Figure 5. Logistic regression (Elastic Net penalty) classifier, trained on US data and tested on Bangladeshi data as well as a held-out test set of the US data.

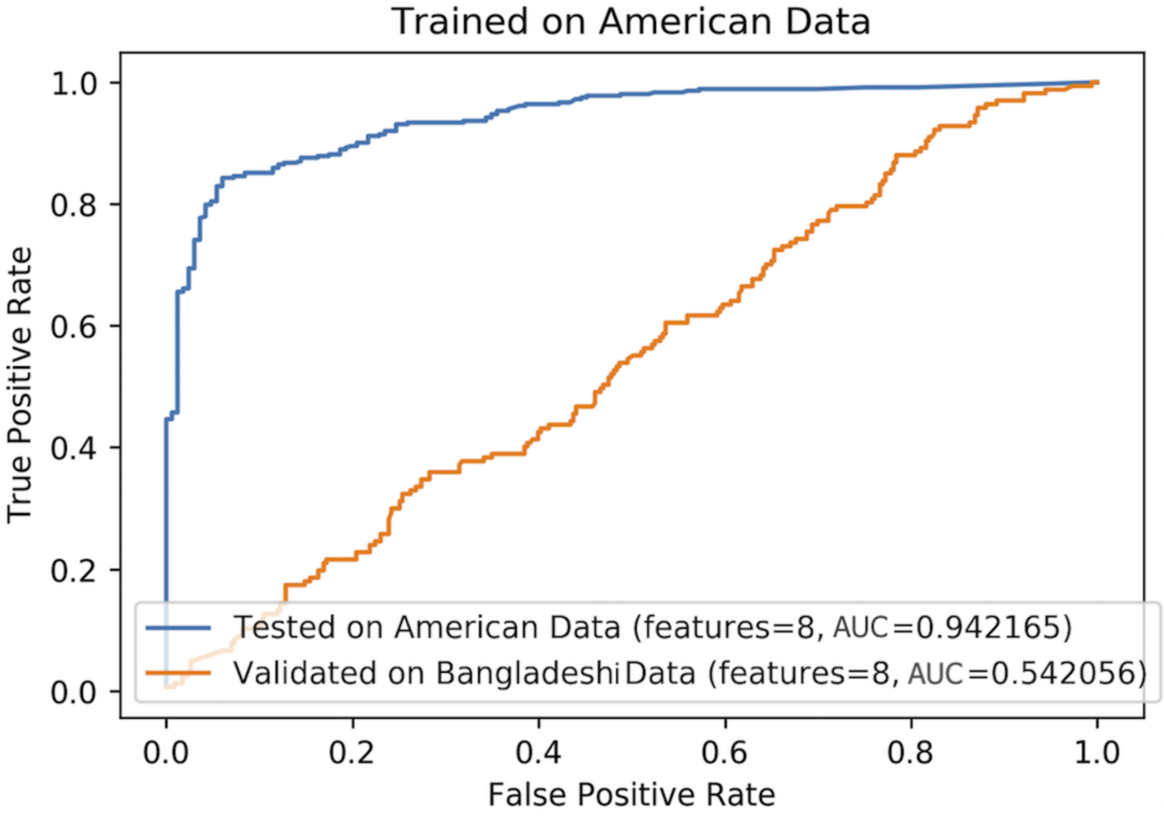
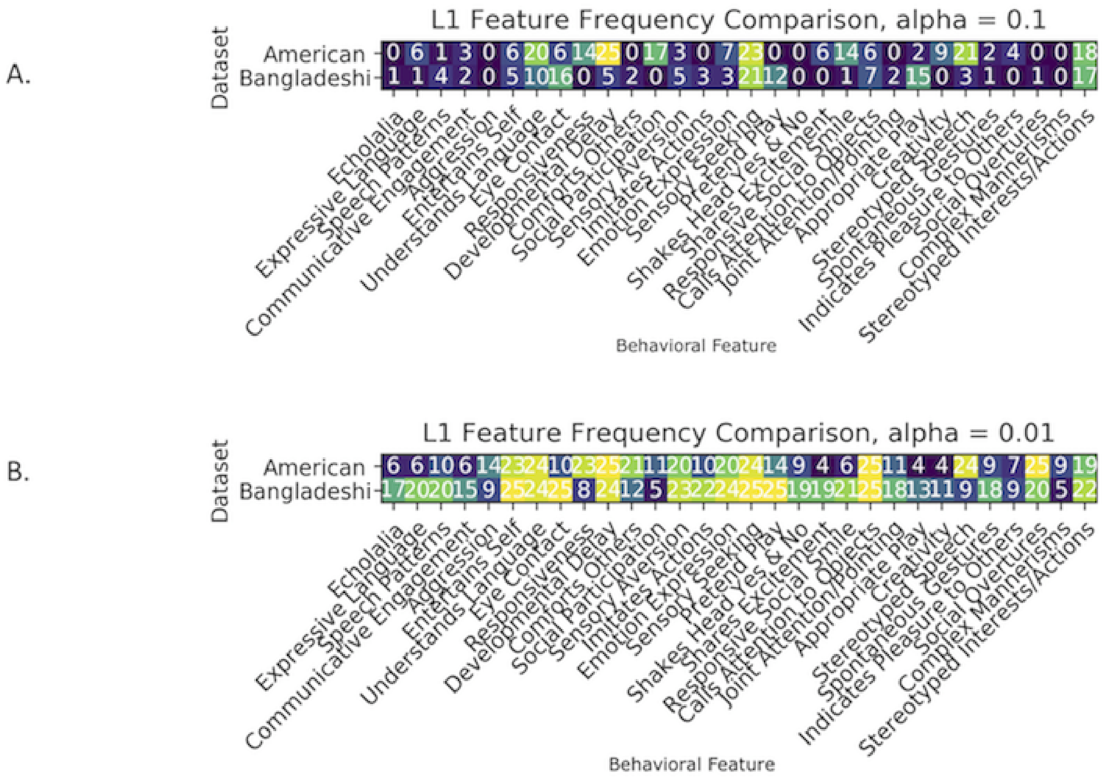


Figure 6. Feature selection analysis. Numbers within the cells indicate the frequency of selection. (A) Feature frequency comparison during cross-fold validation with alpha value 0.1 between Bangladeshi data and US data. (B) Feature frequency comparison during cross-fold validation with alpha value 0.01 between Bangladeshi data and US data.



Discussion

Principal Results

We were able to demonstrate the potential to use video-based machine learning methods to detect developmental delay and autism in a collection of videos of Bangladeshi children at risk for autism. Despite language, cultural, and geographic barriers, this outcome shows promise for remote autism detection in a developing country. More testing and refinement will be needed, but, in general, there is potential for the method to be made virtual to run entirely on mobile devices and therefore potential to increase the capacity to detect and provide more immediate diagnostics to children in need of therapeutic interventions.

An important result of our work is that we were able to gather 159 videos from Bangladeshi parents collected via mobile phone through our collaboration with DSH. This suggests feasibility of expanding this study to a larger sample size across Bangladesh and other low-resource settings and the ability to rely on the use of mobile phones in developing countries like Bangladesh, where 95% of the population are mobile phone subscribers [25]. Additionally, we found that clinically untrained, US-based, non-Bengali speaking raters were able to score videos of Bangladeshi children with limited training, suggesting that speaking the native language may not be necessary for scoring videos. This finding also demonstrates the validity and potential of this mobile tool to be deployed across cultures and languages.

A useful and novel contribution of our work was our method for ensembling predictions from models trained on and adapted to each individual rater. This method demonstrates several advantageous properties. First, because each classification model was trained to map an individual rater's annotation patterns to a predicted class label, these rater-adaptive models can capitalize on features reflecting a rater's strengths while ignoring features on which the rater shows weaker performance. Furthermore, the fact that raters' models are trained independently from one another means that, in a distributed setting where there is a large corpora of videos such that each rater annotates only a small subset of them, our method can make predictions on each video by applying and ensembling the models from each rater without any need for additional imputation. By weighting each rater's model according to its accuracy on a rater-specific held-out validation set, the overall ensemble can lean more heavily on those raters whose models consistently demonstrate the best classification performance. Finally, because the final ensemble's prediction is a linear combination of all of the rater's models and we are able to calculate Shapley values for every feature in each of these models, it follows that we can use the same weights from the ensemble of rater-specific predictions to generate ensemble-level Shapley values as well. Thus, if a child's video is distributed to several different raters and those raters' annotations are fed into the ensemble model, one can interpret how each of the child's behavioral annotations contributed to both the final ensemble classification label and each rater's predicted label individually.

We found that while models trained on videos of US children and models trained on Bangladeshi children both relied on many of the same clinically relevant features (eg, sensory seeking, stereotyped interests, and actions), some features were more prominent in one model compared to the other. For example, models trained on US data tended to rely more heavily on social participation and stereotyped speech, while models trained on Bangladeshi data relied more on eye contact. These patterns make sense, as raters could rely on a mutual understanding of the language (English) to evaluate behaviors like stereotyped speech and social interaction in US videos and may not have needed to rely as heavily on physical cues like eye contact, whereas when US raters viewed Bangladeshi videos, nonlanguage-based cues became more important. Even without the ability to confidently evaluate all aspects of the child's behavior, the rater ensemble demonstrated that the set of behavioral features needed to make an accurate diagnosis of developmental delays, including ASD, may be narrower than previously thought. Nevertheless, the difficulty in assessing certain sociolinguistic patterns in the cross-cultural context may have been the cause of comparatively lower performance in the Bangladeshi dataset. We hypothesize that, when trained on annotations provided by raters who share a common linguistic and sociocultural background with the Bangladeshi children, our ensemble's performance will improve and become comparable to the models trained and evaluated on the US dataset.

Limitations

Although accuracy achieved using our source classifiers originally trained on US datasets was lower when applied to Bangladeshi videos, it still indicated a signal in the Bangladeshi dataset. The relatively low accuracy is most likely a result of three factors. First, these original classifiers were trained on clinical scoresheets, not on features obtained from live video data. Second, these scoresheets were obtained from formal clinical assessments of US children, and therefore they do not capture a culturally diverse set of behavioral nuances. Third, these classifiers were trained to distinguish between typically developing children and children with autism. However, this dataset consists of delays other than autism (eg, SLCs), which may be why these classifiers were unable to classify these cases with higher accuracy.

Conclusions

Although the potential uses for a method of crowdsourced annotation and classification of developmental disorders like the one we established in this work are myriad, we wish to highlight a few uses. First, in areas where resources are scarce, and with a disorder like ASD, where early intervention is the key to successful treatment, our framework could be essential in performing cost-effective and reliable triage. Parents could send short home videos of their children to the cloud, at which point the video would be routed to several raters who perform feature tagging of the child's behavior. Based on the raters' previous annotation patterns and their associated models, the child would receive a predicted risk probability of developmental delay or ASD and a clinical team nearby could then be alerted, as appropriate. Since 2008, Dr Khan and her team have assisted the government to establish multidisciplinary Child Development Centers in tertiary hospitals across Bangladesh [26]. Fifteen such Child Development Centers are currently operational, whose chief mandate is to diagnose and provide appropriate management for a range of neurodevelopmental disorders including autism. However, in a country with population of 160 million, of whom an estimated 45% are within the pediatric age groups, access to reliable services can be limited. Formalization of the approaches documented here could enable broader reach and coverage through remote care while allowing resource-strapped clinical teams to deploy their efforts where they are needed the most.

An exciting second consequence of a deployment like this would be the steady development of a large corpus of annotated videos. No such dataset exists to date; however, the potential impact of such a dataset could be substantial. Modern algorithms from machine vision and speech recognition like convolutional and recurrent neural networks could use these annotations to *learn* features from the raw video and audio that are important for detecting developmental disorders, including ASD. Once trained, these models would dramatically accelerate the speed for detection of disorders and ability to accelerate the delivery of useful interventions.

Another important effect of such a pipeline would be that, with location-tagged videos, we could develop more accurate epidemiological statistics on the prevalence and onset of developmental disorders like ASD worldwide. Better information like this may increase awareness, positively impact policy change, and advance progress for addressing unmet needs of the children with developmental delays. This can have important applications in the developing world by helping countries identify the proportion of the population affected by such delays or impairments and therefore inform policy and gather actionable insights for health sector responses.

Acknowledgments

We are grateful for the generous support and participation of all 159 families who provided video and other phenotypic records. This work was supported by the Bill and Melinda Gates Foundation. It was also supported, in part, by funds to DW from the NIH (1R01EB025025-01 and 1R21HD091500-01), The Hartwell Foundation, the Coulter Foundation, Lucile Packard Foundation, and program grants from Stanford's Human Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics

Center (PHIND), Beckman Center, Bio-X Center, Predictives and Diagnostics Accelerator (SPADA) Spectrum, and Wu Tsai Neurosciences Institute Neuroscience: Translate Program. We also acknowledge the generous support from Peter Sullivan.

Authors' Contributions

QT: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing (original draft, review, and editing). SLF: formal analysis, investigation, methodology, software, validation, visualization, writing (review and editing). JNS: data curation, investigation, methodology, project administration, resources, and writing (review and editing). KD: data curation, investigation, methodology, project administration, resources, writing (review and editing). CC: formal analysis, investigation, methodology, software, writing (review and editing). PW: data curation, formal analysis, resources, writing (review and editing). HK: data curation, formal analysis, resources, writing (review and editing). NZK: conceptualization, funding acquisition, investigation, data curation, methodology, writing (review and editing). GLD: conceptualization, funding acquisition, data curation, writing (review and editing). DPW: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing (original draft, review, and editing)

Conflicts of Interest

DPW is the founder of Cognoa.com. This company is developing digital health solutions for pediatric care. All other authors declare no competing interests.

Multimedia Appendix 1

Formulae used in creating stacked rater-weighted classifiers.

[DOCX File, 116KB - jmir_v21i4e13822_app1.docx]

References

1. Cromer J. U.S. Army. 2018 Apr 05. Autism: fastest-growing developmental disability URL: https://www.army.mil/article/203386/autism_fastest_growing_developmental_disability [accessed 2019-04-17] [WebCite Cache ID 77hgjUBE4]
2. Baio J, Wiggins J, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ 2018 Dec 27;67(6):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6706a1] [Medline: 29701730]
3. Hossain M, Ahmed HU, Jalal Uddin MM, Chowdhury WA, Iqbal MS, Kabir RI, et al. Autism Spectrum disorders (ASD) in South Asia: a systematic review. BMC Psychiatry 2017 Dec 01;17(1):281 [FREE Full text] [doi: 10.1186/s12888-017-1440-x] [Medline: 28826398]
4. Durkin M, Elsabbagh M, Barbaro J, Gladstone M, Happe F, Hoekstra RA, et al. Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. Autism Res 2015 Oct;8(5):473-476 [FREE Full text] [doi: 10.1002/aur.1575] [Medline: 26437907]
5. Khan N, Muslima H, Shilpi AB, Begum D, Parveen M, Akter N, et al. Validation of rapid neurodevelopmental assessment for 2- to 5-year-old children in Bangladesh. Pediatrics 2013 Feb;131(2):e486-e494. [doi: 10.1542/peds.2011-2421] [Medline: 23359579]
6. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord 1989 Jun;19(2):185-212. [Medline: 2745388]
7. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994 Oct;24(5):659-685. [Medline: 7814313]
8. Duda M, Daniels J, Wall DP. Clinical Evaluation of a Novel and Mobile Autism Risk Assessment. J Autism Dev Disord 2016 Dec;46(6):1953-1961 [FREE Full text] [doi: 10.1007/s10803-016-2718-4] [Medline: 26873142]
9. Duda M, Haber N, Daniels J, Wall DP. Crowdsourced validation of a machine-learning classification system for autism and ADHD. Transl Psychiatry 2017 Dec 16;7(5):e1133 [FREE Full text] [doi: 10.1038/tp.2017.86] [Medline: 28509905]
10. Duda M, Kosmicki JA, Wall DP. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. Transl Psychiatry 2014 Aug 12;4:e424 [FREE Full text] [doi: 10.1038/tp.2014.65] [Medline: 25116834]
11. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. Transl Psychiatry 2016 Feb 09;6:e732 [FREE Full text] [doi: 10.1038/tp.2015.221] [Medline: 26859815]
12. Kosmicki J, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 2015 Feb 24;5:e514 [FREE Full text] [doi: 10.1038/tp.2015.7] [Medline: 25710120]

13. Levy S, Duda M, Haber N, Wall DP. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. Mol Autism 2017;8:65 [FREE Full text] [doi: 10.1186/s13229-017-0180-6] [Medline: 29270283]
14. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. PLoS Med 2018 Dec;15(11):e1002705 [FREE Full text] [doi: 10.1371/journal.pmed.1002705] [Medline: 30481180]
15. Wall D, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2012 Apr 10;2:e100 [FREE Full text] [doi: 10.1038/tp.2012.10] [Medline: 22832900]
16. Robins DL, Dumont-mathieu TM. Early Screening for Autism Spectrum Disorders. Journal of Developmental & Behavioral Pediatrics 2006;27(Supplement 2):S111-S119. [doi: 10.1097/00004703-200604002-00009]
17. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 2017:4565-4774.
18. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005;67(2):301-320.
19. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data Mining, Inference, and Prediction. New York: Springer; 2001:101-179.
20. McKinney W. Pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing 2011:14.
21. Pedregosa F. Journal of machine learning research. 2011. Scikit-learn: Machine learning in Python URL: https://dl.acm.org/citation.cfm?id=2078195&preflayout=tabs [accessed 2019-04-17]
22. Bergstra J, Yamins D, Cox DD. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. 2013 Presented at: Proceedings of the 12th Python in science conference; 2013; Austin, Texas.
23. Lundberg S, Erion GG, Lee SI. Cornell University - arXiv.org. 2018. Consistent individualized feature attribution for tree ensembles URL: https://arxiv.org/abs/1802.03888 [accessed 2019-04-17] [WebCite Cache ID 77hkRoF28]
24. Hunter J. Matplotlib: A 2D Graphics Environment. Comput Sci Eng 2007 May;9(3):90-95. [doi: 10.1109/MCSE.2007.55]
25. BTRC. 2018. Mobile Phone Subscribers in Bangladesh December, 2018 URL: http://www.btrc.gov.bd/content/mobile-phone-subscribers-bangladesh-december-2018 [accessed 2019-04-17] [WebCite Cache ID 77hl4ohnk]
26. Khan N, Sultana R, Ahmed F, Shilpi AB, Sultana N, Darmstadt GL. Scaling up child development centres in Bangladesh. Child Care Health Dev 2018 Dec;44(1):19-30. [doi: 10.1111/cch.12530] [Medline: 29235172]

Abbreviations

ADI-R: Autism Diagnostic Interview-Revised
ADOS: Autism Diagnostic Observation Schedule
ASD: autism spectrum disorder
AUC: area under the curve
DSH: Dhaka Shishu Children’s Hospital
HIPAA: Health Insurance Portability and Accountability Act
MCHAT: Modified Checklist for Autism in Toddlers
SLC: speech and language condition
TD: neurotypical development

Edited by G Eysenbach; submitted 28.02.19; peer-reviewed by E Linstead, S Teraman; comments to author 20.03.19; revised version received 03.04.19; accepted 05.04.19; published 24.04.19.

Please cite as:
Tariq Q, Fleming SL, Schwartz JN, Dunlap K, Corbin C, Washington P, Kalantarian H, Khan NZ, Darmstadt GL, Wall DP
Detecting Developmental Delay and Autism Through Machine Learning Models Using Home Videos of Bangladeshi Children: Development and Validation Study
J Med Internet Res 2019;21(4):e13822
URL: http://www.jmir.org/2019/4/e13822/
doi:10.2196/13822
PMID:31017583

Original Paper

Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey

Adam Palanica[1], PhD; Peter Flaschner[2], BDES; Anirudh Thommandram[1], MAS; Michael Li[1], MAS; Yan Fossat[1], CAP

[1]Labs Department, Klick Health, Klick Inc, Toronto, ON, Canada
[2]Marketing Department, Klick Health, Klick Inc, Toronto, ON, Canada

Corresponding Author:
Adam Palanica, PhD
Labs Department
Klick Health
Klick Inc
175 Bloor St E, Suite 300
Toronto, ON, M4W 3R8
Canada
Phone: 1 416 214 4977
Email: apalanica@klick.com

Abstract

Background: Many potential benefits for the uses of chatbots within the context of health care have been theorized, such as improved patient education and treatment compliance. However, little is known about the perspectives of practicing medical physicians on the use of chatbots in health care, even though these individuals are the traditional benchmark of proper patient care.

Objective: This study aimed to investigate the perceptions of physicians regarding the use of health care chatbots, including their benefits, challenges, and risks to patients.

Methods: A total of 100 practicing physicians across the United States completed a Web-based, self-report survey to examine their opinions of chatbot technology in health care. Descriptive statistics and frequencies were used to examine the characteristics of participants.

Results: A wide variety of positive and negative perspectives were reported on the use of health care chatbots, including the importance to patients for managing their own health and the benefits on physical, psychological, and behavioral health outcomes. More consistent agreement occurred with regard to administrative benefits associated with chatbots; many physicians believed that chatbots would be most beneficial for scheduling doctor appointments (78%, 78/100), locating health clinics (76%, 76/100), or providing medication information (71%, 71/100). Conversely, many physicians believed that chatbots cannot effectively care for all of the patients' needs (76%, 76/100), cannot display human emotion (72%, 72/100), and cannot provide detailed diagnosis and treatment because of not knowing all of the personal factors associated with the patient (71%, 71/100). Many physicians also stated that health care chatbots could be a risk to patients if they self-diagnose too often (714%, 74/100) and do not accurately understand the diagnoses (74%, 74/100).

Conclusions: Physicians believed in both costs and benefits associated with chatbots, depending on the logistics and specific roles of the technology. Chatbots may have a beneficial role to play in health care to support, motivate, and coach patients as well as for streamlining organizational tasks; in essence, chatbots could become a surrogate for nonmedical caregivers. However, concerns remain on the inability of chatbots to comprehend the emotional state of humans as well as in areas where expert medical knowledge and intelligence is required.

(J Med Internet Res 2019;21(4):e12887) doi:10.2196/12887

KEYWORDS

physician satisfaction; health care; telemedicine; mobile health; health surveys

Introduction

Background

Chatbots, also known as conversational agents, interactive agents, virtual agents, virtual humans, or virtual assistants, are artificial intelligence programs designed to simulate human conversation via text or speech. Many positive viewpoints have been made on the potential uses of health care chatbots within the marketing and business world [1-10]; however, little scientific research has examined their effectiveness in real-world patient scenarios, that is, to improve health outcomes [11,12]. Chatbots are commonly used in marketing applications such as to guide consumers through electronic commerce websites, answer questions related to products and services, help troubleshoot problems with internet service, act as a personal concierge, or provide consumer advice. In the context of health care, chatbots or *healthbots* are intended to provide personalized health and therapy information to patients, provide relevant products and services to patients, as well as suggest diagnoses and recommend treatments based on patient symptoms.

Chatbots in health care may have the potential to provide patients with access to immediate medical information, recommend diagnoses at the first sign of illness, or connect patients with suitable health care providers (HCPs) across their community [13,14]. Theoretically, in some instances, chatbots may be better suited to help patient needs than a human physician because they have no biological gender, age, or race and elicit no bias toward patient demographics. Chatbots do not get tired, fatigued, or sick, and they do not need to sleep; they are cost-effective to operate and can run 24 hours a day, which is especially useful for patients who may have medical concerns outside of their doctor’s operating hours. Chatbots can also communicate in multiple different languages to better suit the needs of individual patients.

Early research has demonstrated the benefits of using health care chatbots, such as helping with diagnostic decision support [15,16], promoting and increasing physical activity [17], and cognitive behavioral therapy for psychiatric and somatic disorders [18-24], which provide effective, acceptable, and practical health care with accuracy comparable with that of human physicians. Patients may also feel that chatbots are *safer* interaction partners than human physicians and are willing to disclose more medical information and report more symptoms to chatbots [25,26]. However, despite the demonstrated efficacy and cost-effectiveness of health care chatbots, the technology is usually associated with poor adoption by physicians and poor adherence by patients [27]. This may be because of the perceived lack of quality or accountability that is characterized by computerized chatbots as opposed to traditional face-to-face interactions with human physicians.

Objectives

Although there are some instances showing the effectiveness of health care–related chatbots for certain outcomes, it is still not entirely clear whether this technology is better overall in improving all the various clinical health outcomes of patients and why it is not more highly adopted compared with traditional methods of care, that is, information coming from a human physician. Although chatbot technology for health care is continually advancing, little is known about the perspectives of practicing medical physicians on the use of chatbots in health care. It would thus seem beneficial to have medical expert opinions on the use of this technology that is intended to supplement or even replace specific roles of HCPs. The purpose of this study was to examine the perspectives of practicing medical physicians on the use of health care chatbots for patients. As human physicians have been the traditional benchmark for treating patients for hundreds of years, a crucial objective of investigating the use of chatbots for delivering health care should be to understand the perspective of medical experts who actually practice health care in their daily occupations. As physicians are the primary point of care for patients, their approval is an important gate to the dissemination of chatbots into medical practice. The findings of this research will help to either justify or attenuate enthusiasm for health care chatbot applications as well as direct future work to better align with the needs of HCPs.

Methods

Participants

A total of 100 participants completed the survey (28 females, 69 males, and 3 preferred not to say; age range=28-73 years; mean age 44.9, SD 12.0). Participants were general practitioners (GPs) with a Doctor of Medicine (MD) degree (years of practice range=2-36; mean years of practice 14.7, SD 9.0). Participants were located across 32 states of the United States, with participants in each of the 4 main interstate regions of Northeast (27/100, 27%; 5 states), Midwest (21/200, 21%; 8 states), South (29/100, 29%; 13 states), and West (23/100, 23%; 6 states).

Recruitment

Participants who took part in the survey were sampled from a large database of physicians who have previously agreed to take part in market research. The survey was administered by Sermo [28], a private social media network for licensed physicians, who randomly selected registered physicians within their panel across the United States. The Sermo research network comprises over 400,000 registered physicians in the United States, representing roughly 40% of the US physician population [19]. As this study was the first of its kind and exploratory in nature to study the subjective opinions of physicians, no explicit statistical hypotheses were being evaluated. The sample size of 100 was arbitrarily chosen to gather a preliminary viewpoint of physicians’ perspectives of chatbots in health care and would yield approximately a 9.8% margin of error with a 95% CI of the entire US physician population.

Invitees were sent an email inviting them to complete the survey, accessible via an embedded Web link. The only inclusion criteria included being a GP with an MD degree within the United States; no restrictions on age, gender, or previous use of chatbots were implemented. Exclusion criteria included any nonphysician or specialist physician, which may have had a bias toward the use of health care chatbots for patients; in contrast, GPs are typically the primary point of care for a broad range of family medicine–related health issues, so these participants were targeted for the research. All participants gave informed consent to complete the survey, and the study received full ethics clearance from Advarra Institutional Review Board Services, an independent ethics committee.

Survey and Procedure

Survey questions were designed in consultation with medical scientists, Web developers, data scientists, and technology specialists with expertise in digital medicine. The survey was organized into 5 main sections, including (1) usage of health care chatbots, (2) perceived benefits of health care chatbots to patients, (3) perceived challenges of health care chatbot usage, (4) perceived risks of health care chatbots, and (5) physicians' perceptions of health care chatbots in the role of a physician.

Participants were asked to take part in a research study involving a Web-based self-report survey that examines physician perceptions of the benefits, challenges, and risks of using chatbots in health care. Participants were asked to provide demographic information and report their opinions on the use of chatbot technology for treating patients in health care. At the beginning of the survey, participants were given an explicit definition of *chatbot*, to give a baseline description for the context of the questions in the survey. Participants were given the following definition:

> *Chatbots, also known as conversational agents, interactive agents, virtual agents, virtual humans, or virtual assistants, are computer software applications that run automated tasks or scripts designed to simulate human conversation. Chatbots are artificial intelligence (AI) programs that can generate and retrieve information for the interaction with human users via text or computer voice generation.*

Participants were asked to answer all the survey questions for chatbots in the context of health care, referring to the use of chatbots for health-related issues.

Data Analysis

Data were analyzed using descriptive statistics and frequencies to examine the characteristics of participant responses to survey items on health care chatbots. Preliminary analyses revealed no major differences across factors of age, gender, or years of practice. Thus, the entire sample was reported holistically.

Results

Usage of Health Care Chatbots

A total of 30% (30/100) of participants indicated that they had direct personal experience with the use of chatbots for health-related issues. Physicians were also given a list of currently available health care chatbots, to examine their familiarity with some of the interfaces that could be potentially accessed by patients. Table 1 shows physicians' use of these health care chatbots, which are intended to provide personalized health and therapy information, provide relevant products and services to patients, as well as suggest diagnoses and recommend treatments based on patients' symptoms. The findings indicated that most of the currently available chatbots were not generally used or heard of by physicians.

Of the 30 participants who have used health care chatbots previously, 4 (13%) were very satisfied, 10 (33%) were somewhat satisfied, 8 (27%) were neither satisfied nor dissatisfied, and 8 (27%) were somewhat dissatisfied with their application. Of all the physicians in the survey, 18% (18/100) stated that their patients use health care chatbots (24%, 24/100, stated that patients did not use them), but the majority (58%, 58/100) were unsure or did not know whether their patients use them.

In total, 42% (42/100) of physicians believed that chatbots are either very important (9%, 9/100) or somewhat important (33%, 33/100) in health care, whereas 26% (26/100) believed that they are somewhat unimportant (18%, 18/100, 18%) or very unimportant (8%, 8/100); 32% (32/100) of physicians believed that they are neither important nor unimportant. Similarly, 44% (44/100) of physicians stated that they would be very likely (9/100, 9%) or somewhat likely (35%, 35/100) to prescribe the use of health care chatbots to their patients within the next 5 years; 34% (34/100) of physicians stated that they would be somewhat unlikely (22/100, 22%) or very unlikely (12/100, 12%) to do so. A total of 40% (40/100) of physicians also indicated that they would be very likely (11/100, 11%) or somewhat likely (29%, 29/100) to recommend the prescription of health care chatbots to their HCP colleagues, whereas 37% (37/100) indicated that they would be somewhat unlikely (25%, 25/100) or very unlikely (12%, 12/100) to do the same.

Perceived Benefits of Health Care Chatbots to Patients

Participants were asked to what extent they thought health care chatbots would benefit patients in specific areas of health management (Table 2). An average of 42% (42/100) agreed to some extent in the benefits associated with health care chatbots, whereas an average of 25% (25/100) disagreed to some extent in these same potential benefits. More than half of physicians agreed that health care chatbots could help patients better manage their own health (54/100, 54%), improve access and timeliness to care (53%, 53/100), or reduce travel time to their HCP (52%, 52/100); almost half of physicians believed that health care chatbots could prevent unnecessary visits to HCPs (49/100, 49%) or that patients may disclose more information to chatbots compared with HCPs (41%, 41/100).

In terms of specific health-related outcomes of chatbot use for patients, an average of 45% (45/100) of physicians believed in some type of physical, psychological, or behavioral health benefit to patients (Table 3). More than half of physicians believed that health care chatbots could improve nutrition or diet (65%, 65/100), enhance medication or treatment adherence (60%, 60/100), increase activity or exercise (55%, 55/100), or reduce stress (51%, 51/100).

Table 1. Participants' use of currently available health care chatbots.

Currently available health care chatbots[a]	Used it, %	Heard of it but never used it, %	Neither heard of it nor used it, %
Your.MD	14	21	65
HealthTap	11	17	72
Cancer Chatbot	8	13	79
VitaminBot	7	13	80
Babylon Health	5	14	81
Safedrugbot	5	10	85
Ada Health	3	12	85
Bots4Health	3	14	83
SimSensei	3	12	85
Sensely	2	12	86
Infermedica	2	16	82
Izzy	2	11	87
Luma Health	2	14	84
Ginger.io	2	12	86
Buoy Health	1	14	85
Florence	1	9	90
HelloJoy	1	11	88
Woebot	0	13	87
GYANT	0	9	91

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Table 2. Perceived benefits of health care chatbots to patients.

Perceived benefits of health care chatbots[a]	Strongly agree, %	Somewhat agree, %	Neither agree nor disagree, %	Somewhat disagree, %	Strongly disagree, %
Help patients better manage their own health	7	47	26	17	3
Improve quality of patient care	8	27	36	21	8
Help provide more personalized treatment	7	21	36	27	9
Reduce travel time to health care provider	20	32	30	12	6
Prevent unnecessary visits to health care providers	11	38	28	19	4
Patients may disclose more information to chatbots compared with health care providers	12	29	36	15	8
Increase patient privacy	8	19	40	20	13
Improve access and timeliness to care	15	38	32	12	3
Average across variables	11	31	33	18	7

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Table 3. Perceived health care outcome benefits of using chatbots for patients.

Perceived health care outcome benefits of using chatbots[a]	Yes, %	No, %	Do not know or not sure, %
Activity or exercise increase	55	25	20
Alcohol consumption reduction	31	36	33
Blood glucose reduction	49	26	25
Blood pressure reduction	34	32	34
Cholesterol reduction	32	37	31
Cognitive behavioral therapy	41	30	29
Depression reduction	33	37	30
Medication/treatment adherence	60	17	23
Nutrition/diet improvement	65	19	16
Sleep quality or quantity improvement	34	35	31
Smoking reduction or cessation	47	24	29
Stress reduction or management	51	23	26
Psychological well-being increase	48	28	24
Weight loss or decrease in body mass index	45	30	25
Average across variables	45	29	27

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Table 4. Perceived logistical benefits of using chatbots for patients.

Perceived logistical benefits of using chatbots[a]	Yes, %	No, %	Do not know or not sure, %
Locating health clinics or health care providers in a specific area	76	11	13
Scheduling doctor appointments	78	13	9
Monitoring patient calls to the reception desk of health clinics	49	22	29
Processing medical invoices or bill payments	48	28	24
Assessing emergency triage in hospitals	30	48	22
Reminders for medication/treatment compliance	76	11	13
Renewing medication prescriptions	56	25	19
Gathering health insurance information	65	16	19
Answering medication frequently asked questions	70	15	15
Providing medication side effects and drug interactions	68	15	17
Providing medication use or misuse instructions	71	12	17
Average across variables	62	20	18

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Regarding logistical benefits of using health care chatbots for patients, the majority of physicians (62%, 62/100) believed in advantages for organization, planning, and management of administrative characteristics associated with health care (Table 4). Most notably, many physicians believed that chatbots would be most beneficial for scheduling doctor appointments (78%, 78/100), locating health clinics or HCPs in a specific area (76%, 76/100), administering reminders for medication or treatment compliance (76%, 76/100), providing medication use or misuse instructions (71%, 71/100), or answering medication frequently asked questions (70%, 70/100). In contrast, almost half of physicians believed that chatbots could *not* assess emergency triage in hospitals (48%, 48/100).

Perceived Challenges Associated With Health Care Chatbots for Patients

Over half of the physicians (53%, 53/100) agreed to some extent in various challenges associated with the use of health care chatbots for patients, whereas an average of only 14% (14/100) disagreed with these traits (Table 5). Most notably, 76% (76/100) of physicians believed that chatbots cannot effectively care to the full extent of the patients' needs, 72% (72/100) believed that chatbots cannot understand or display human emotion, and 58% (58/100) believed that chatbots lack the intelligence or knowledge to accurately assess patients.

Table 5. Perceived challenges associated with using health care chatbots for patients.

Perceived challenges associated with using health care chatbots[a]	Strongly agree, %	Somewhat agree, %	Neither agree nor disagree, %	Somewhat disagree, %	Strongly disagree, %
Patient data privacy and confidentiality	17	31	39	11	2
Chatbots cannot understand or display human emotion	36	36	23	4	1
Chatbots lack the intelligence or knowledge to accurately assess patients	19	39	27	11	4
Chatbots offer poor health-related advice	12	28	47	13	0
Chatbots cannot effectively care to the full extent of the patients' needs	43	33	17	7	0
Chatbots take too much time to use	6	24	48	20	2
Most of my patients do not have access to the necessary technology for chatbots services	15	35	26	18	6
Average across variables	21	32	32	12	2

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Table 6. Perceived risks associated with using health care chatbots for patients.

Perceived risks associated with using health care chatbots	Strongly agree, %	Somewhat agree, %	Neither agree nor disagree, %	Somewhat disagree, %	Strongly disagree, %
Patients may abuse the use of chatbots and self-diagnose too often	21	53	20	5	1
Patients will receive lesser quality assessments	24	38	28	8	2
Patients may not accurately understand the diagnoses	26	48	22	2	2
Chatbots cannot provide detailed clarification on patient assessment	35	36	19	8	2
Patients may not feel adequately connected to their health care providers	34	36	22	8	0
Chatbots may indirectly harm patients by not knowing all of the personal factors associated with the patient	28	41	24	6	1
Average across variables	28	42	23	6	1

[a] For simplicity, as there were exactly 100 participants in the study, only percentages have been reported, unless otherwise stated.

Perceived Risks Associated With Health Care Chatbots for Patients

The great majority of physicians (70%, 70/100) expressed their concern of risks associated with health care chatbots for patients, whereas only 7% (7/100) disagreed with these potential risks (Table 6). Over 60% (60/100) of physicians agreed to every type of risk presented to them, including the perception that patients may abuse the use of chatbots and self-diagnose too often (74%, 74/100), patients may not accurately understand the diagnoses (74%, 74/100), and that chatbots cannot provide detailed clarification on patient assessment (71%, 71/100). Physicians also felt that patients may not feel adequately connected to their HCPs (70%, 70/100) or that chatbots may indirectly harm patients by not knowing all of the personal factors associated with the patient (69%, 69/100).

Physicians' Perceptions of Health Care Chatbots in the Role of a Physician

Physicians were asked how much they believed that health care chatbots would either help them or impede their work in their daily occupational role on a sliding scale from 0%=*impede my work* to 100%=*help me*. Physicians responded entirely across the board, which averaged at an approximate neutral point in the middle of the scale (observed range=0-100%, mean 47.4%, median 50%, SD 25.6%).

Finally, physicians were asked how likely it would be, in the future, for health care chatbots to play a more significant role in patients' health than their HCP. A total of 49% (49/100) expressed that this would be very likely (15%, 15/100) or somewhat likely (34%, 34/100) to happen, whereas 25%

(25/100) expressed that this would be somewhat unlikely (15%, 15/100) or very unlikely (10%, 10/100) to happen.

Discussion

Principal Findings

A total of 100 practicing GPs participated in an online research survey that examined their perceived benefits, challenges, and risks of using chatbots in health care. Overall, the findings demonstrated that physicians have a wide variety of perspectives on the use of health care chatbots for patients, with few major skews to one side or the other regarding agreement levels to a variety of characteristics. Almost half of the physicians perceived health care chatbots to be important for patients, especially for helping patients better manage their own health. Almost half of the physicians also stated that they would be likely to prescribe the use of the technology to patients and recommend it to their colleagues. About half of the physicians also agreed that chatbots would benefit the physical, psychological, and behavioral health outcomes of patients, such as diet improvement, medication adherence, exercise frequency, or stress reduction. The other half of physicians was roughly equally divided between being an opponent or having a neutral opinion to the perceived importance and benefits of health care chatbots. In addition, patient privacy did not emerge as a polarizing issue.

With regard to the use of health care chatbots within the occupational role of an HCP, physicians believed that the technology would almost equally help them as well as impede their overall workplace duties. Approximately half of the physicians also believed that health care chatbots would eventually play a more significant role in patients' health than their HCP. For the most part, these results indicated an almost equal number of supporters for health care chatbots, with the rest being those who are either indifferent or opponents to the technology.

More consistent agreement on the use of health care chatbots was apparent with reference to their potential logistical benefits as well as their challenges and potential risks to patients. For example, the great majority of physicians believed in administrative benefits associated with chatbots, especially for scheduling doctor appointments, locating health clinics or HCPs, administering reminders for medication compliance, providing treatment instructions, and answering commonly asked medication questions. In contrast, the majority of physicians believed that chatbots cannot effectively care for all of the patients' needs, cannot understand or display human emotion, lack the intelligence to accurately assess patients, cannot provide detailed clarification on patient assessment, cannot assess emergency health situations, or may indirectly harm patients by not knowing all of the personal factors associated with the patient. In addition, many physicians stated that health care chatbots will be associated with the risk that patients may self-diagnose too often, patients may not understand the diagnoses, or that patients may not feel adequately connected to their primary physician. These findings highlight the perception that chatbot technology may be advantageous to use in less complicated roles, such as administrative and organizational tasks, but more challenges and risks may be associated with their use in complex roles that involve more personalized knowledge of the patient. This suggests that health care duties involving an *expert human touch* and those that need a high degree of accuracy are perceived to be a poor choice for chatbot use compared with receiving overall treatment from an actual physician.

The many perceived challenges and risks associated with health care chatbots would need to be addressed before the technology is widely endorsed by practicing physicians. These challenges may be because of concerns involving regulation and remuneration to the physicians, which supports other relevant research demonstrating that physicians are less likely to use telemedicine services if they are not adequately compensated for their time and effort [29-31]. Addressing the perceived barriers around health care chatbots would, therefore, require cooperation by health care institutions, policy makers, HCPs, and patients alike.

Limitations

One limitation to this research is that it examined the subjective perceptions of GPs as opposed to specialist physicians, which may have had more experience and different opinions on the use of health care chatbots, depending on their roles in patient health. In addition, all physicians practiced within the United States, which may be associated with a different level of enthusiasm toward digital medicine technology compared with other countries and cultures. Some research and viewpoints on health care chatbots have been published from international researchers around the world [13-27]; however, to the best of our knowledge, this was the first study to examine physicians' perspectives on the direct use of chatbots in their practice. Future research should examine how different samples of HCPs, in different environments, perceive health care chatbots for use with their patients.

Conclusions

Physicians believe in both costs and benefits associated with chatbots, depending on the logistics and specific roles of the technology. The areas where physicians believed chatbots would be most helpful were in the improvement of nutrition, diet, and treatment compliance as well as logistical tasks such as scheduling appointments, locating clinics, and providing medication reminders. The major challenges perceived were an inability of chatbots to understand emotions and address the full extent of a patient's needs. Physicians also agreed that there were significant risks associated with chatbots including inaccurate medical information. These findings suggest that physicians may be comfortable with using chatbots to automate simple logistical tasks but do not believe that chatbots are advanced enough to replace complex decision-making tasks requiring an expert medical opinion. This is not to say that health care chatbots have a particular stigma associated with them, but rather, this suggests that improvements are needed for future use to overcome the risks and challenges associated with the technology. Nevertheless, nearly half of the physicians believed that health care chatbots could replace a major role of human

HCPs sometime in the future. However, chatbots can be best applied to help physicians rather than replace them. Chatbots are cost-effective to run and can automate repetitive administrative tasks, thus freeing time for physicians to provide higher quality, personalized, and empathetic care to their patients. This research lays the foundation for future investigations on the factors influencing physician adoption of chatbots. Providing physicians with evidence-based research on the advantages and disadvantages of this emerging technology will help inform them on the most appropriate use to complement their practice rather than impede their work.

Acknowledgments

The authors would like to thank all the participants, project members, supporters, and researchers at Klick Inc for the successful development, implementation, and evaluation of this research. The authors would also like to acknowledge Gaurav Baruah and Peter Leimbigler for their helpful comments on the research design and survey. This research was internally funded and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

References

1. Brooke S. Chatbots Magazine. How chatbots will shape the future of healthcare URL: https://chatbotsmagazine.com/how-chatbots-will-shape-the-future-of-healthcare-fa8e30cebb1c?gi=7238e40fd943 [accessed 2019-02-25] [WebCite Cache ID 76SM3rRld]
2. Cade J. Medium. No appointment necessary — The rise of bots in healthcare URL: https://medium.com/gene-global/no-appointment-necessary-the-rise-of-bots-in-healthcare-4d42c5539547 [accessed 2019-02-25] [WebCite Cache ID 76SM9YAvq]
3. Farkash Z. Chatbots Life. Chatbot for healthcare: chatbots can be money-savers for hospitals and clinics URL: https://chatbotslife.com/chatbot-for-healthcare-chatbots-can-be-money-savers-for-hospitals-and-clinics-5c0c942b4a7a?gi=7510b540e6d0 [accessed 2019-02-25] [WebCite Cache ID 76SMCiqSy]
4. Luma Health. Medium. Using bots in healthcare URL: https://medium.com/healthfurther/using-bots-in-healthcare-42f2cd1f5e4b [accessed 2019-02-25] [WebCite Cache ID 76SMFK9ps]
5. Bryant M. Healthcare Dive. Hospitals turn to chatbots, AI for care URL: https://www.healthcaredive.com/news/chatbots-ai-healthcare/516047/ [accessed 2019-02-25] [WebCite Cache ID 76SMHp2f9]
6. Carter S. ScienceSoft. How chatbots and AI are changing the healthcare industry URL: https://www.scnsoft.com/blog/how-chatbots-and-ai-are-changing-the-healthcare-industry [accessed 2019-02-25] [WebCite Cache ID 76SMMjc2B]
7. Garvin E. HIT Consultant. 5 use cases for chatbots in healthcare URL: https://hitconsultant.net/2018/03/21/chatbots-in-healthcare-use-cases/ [accessed 2019-02-25] [WebCite Cache ID 76SMXq88u]
8. ISDDesign. Medium. 5 advantages of chatbots in the healthcare industry URL: https://medium.com/@ISDDesign/5-advantages-of-chatbots-in-the-healthcare-industry-622dd9253384 [accessed 2019-02-25] [WebCite Cache ID 76SMdK2i2]
9. Maruti Techlabs. Chatbots Magazine. Is conversational AI the future of healthcare? URL: https://chatbotsmagazine.com/is-conversational-ai-the-future-of-healthcare-658a3d8e9dd5?gi=80b5350f28f2 [accessed 2019-02-25] [WebCite Cache ID 76SMlufyj]
10. Parayil M. Chatbots Life. How a chatbot can help your healthcare business URL: https://chatbotslife.com/chatbots-for-healthcare-ab364c8b6e0?gi=e1ea2c1a800a [accessed 2019-02-25] [WebCite Cache ID 76SMvOiCY]
11. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]
12. Steinhubl S, Muse E, Topol E. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
13. Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A, Sansone C. CEUR Workshop Proceedings. 2017. Chatbots Meet eHealth: Automatizing Healthcare URL: http://ceur-ws.org/Vol-1982/paper6.pdf [accessed 2019-02-26] [WebCite Cache ID 76TYvkIFI]
14. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
15. Feldman MJ, Hoffer EP, Barnett GO, Kim RJ, Famiglietti KT, Chueh HC. Impact of a computer-based diagnostic decision support tool on the differential diagnoses of medicine residents. J Grad Med Educ 2012 Jun;4(2):227-231 [FREE Full text] [doi: 10.4300/JGME-D-11-00180.1] [Medline: 23730446]

16. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. arXiv. 2018. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis URL: https://arxiv.org/pdf/1806.10698.pdf [accessed 2019-02-26] [WebCite Cache ID 76Ta0km9I]
17. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
18. 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]
19. Chan S, Godwin H, Gonzalez A, Yellowlees PM, Hilty DM. Review of use and integration of mobile apps into psychiatric treatments. Curr Psychiatry Rep 2017 Oct 30;19(12):96. [doi: 10.1007/s11920-017-0848-9] [Medline: 29082425]
20. Fitzpatrick K, 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 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
21. 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]
22. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Dec 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
23. 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]
24. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]
25. Lucas G, Gratch J, King A, Morency L. It’s only a computer: virtual humans increase willingness to disclose. Comp Hum Behav 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
26. Lucas GM, Rizzo A, Gratch J. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI 2017;4:51. [doi: 10.3389/frobt.2017.00051]
27. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
28. Sermo: talk real world medicine. URL: http://www.sermo.com/ [accessed 2019-02-25] [WebCite Cache ID 76SNPAAHe]
29. Pong R, Hogenbirk J. Reimbursing physicians for telehealth practice: issues and policy options. Health Law Rev 2000;9(1):3-13 [FREE Full text]
30. de Grood C, Raissi A, Kwon Y, Santana M. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
31. Anderson K, Francis T, Ibanez-Carrasco F, Globerman J. Physician's perceptions of telemedicine in HIV care provision: a cross-sectional web-based survey. JMIR Public Health Surveill 2017 May 30;3(2):e31 [FREE Full text] [doi: 10.2196/publichealth.6896] [Medline: 28559226]

Abbreviations

GP: general practitioner
HCP: health care provider
MD: Doctor of Medicine

Edited by G Eysenbach; submitted 22.11.18; peer-reviewed by B Davis, R Alkoudmani; comments to author 31.01.19; revised version received 05.02.19; accepted 09.02.19; published 05.04.19.

Please cite as:
Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y
Physicians’ Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey
J Med Internet Res 2019;21(4):e12887
URL: https://www.jmir.org/2019/4/e12887/
doi:10.2196/12887
PMID:30950796

Original Paper

Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study

SoHyun Park[1], MS; Jeewon Choi[2], MA; Sungwoo Lee[1], BS; Changhoon Oh[1], PhD; Changdai Kim[3], EdD; Soohyun La[4], MA; Joonhwan Lee[2], PhD; Bongwon Suh[1], PhD

[1]Human Centered Computing Lab., Seoul National University, Seoul, Republic of Korea
[2]Human Computer Interaction + Design Lab., Seoul National University, Seoul, Republic of Korea
[3]Department of Education, Seoul National University, Seoul, Republic of Korea
[4]Center for Campus Life and Culture, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Bongwon Suh, PhD
Human Centered Computing Lab.
Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 31 888 9128
Email: bongwon@snu.ac.kr

Abstract

Background: In addition to addiction and substance abuse, motivational interviewing (MI) is increasingly being integrated in treating other clinical issues such as mental health problems. Most of the many technological adaptations of MI, however, have focused on delivering the action-oriented treatment, leaving its relational component unexplored or vaguely described. This study intended to design a conversational sequence that considers both technical and relational components of MI for a mental health concern.

Objective: This case study aimed to design a conversational sequence for a brief motivational interview to be delivered by a Web-based text messaging application (*chatbot*) and to investigate its conversational experience with graduate students in their coping with stress.

Methods: A brief conversational sequence was designed with varied combinations of MI skills to follow the 4 processes of MI. A Web-based text messaging application, Bonobot, was built as a research prototype to deliver the sequence in a conversation. A total of 30 full-time graduate students who self-reported stress with regard to their school life were recruited for a survey of demographic information and perceived stress and a semistructured interview. Interviews were transcribed verbatim and analyzed by Braun and Clarke’s thematic method. The themes that reflect the process of, impact of, and needs for the conversational experience are reported.

Results: Participants had a high level of perceived stress (mean 22.5 [SD 5.0]). Our findings included the following themes: Evocative Questions and Clichéd Feedback; Self-Reflection and Potential Consolation; and Need for Information and Contextualized Feedback. Participants particularly favored the relay of evocative questions but were less satisfied with the agent-generated reflective and affirming feedback that filled in-between. Discussing the idea of change was a good means of reflecting on themselves, and some of Bonobot’s encouragements related to graduate school life were appreciated. Participants suggested the conversation provide informational support, as well as more contextualized feedback.

Conclusions: A conversational sequence for a brief motivational interview was presented in this case study. Participant feedback suggests sequencing questions and MI-adherent statements can facilitate a conversation for stress management, which may encourage a chance of self-reflection. More diversified sequences, along with more contextualized feedback, should follow to offer a better conversational experience and to confirm any empirical effect.

(J Med Internet Res 2019;21(4):e12231) doi:10.2196/12231

KEYWORDS

motivational interviewing; mental health; conversational agents; stress management

Introduction

Background

In recent years, there has been a soaring number of technological adaptations of motivational interviewing (MI) [1]. Most of them, however, focus on changing problematic physical health and lifestyle behaviors (eg, [2-14]). This may be due to the fact that MI primarily targets behavior change and was originally introduced to treat substance abuse, such as addiction and drinking problems [15]. However, recent studies include MI in mental health issues, such as anxiety, depression, and other related problems (eg, [16-22]). It is increasingly acknowledged that MI can be used in a broader and more flexible context concerning ambivalence in change [16]. Most recently, MI took the means of life coaching for college students to cope with stress, yielding positive client experiences in stress reduction [23].

Whether MI is used in treating physical or mental health problems, it is stressed that the counsellor maintain a good MI spirit, that is, the counsellor being relationally adept in expressing empathy and responding to client resistance, in addition to the technical approach [24,25]. However, most technological adaptations of MI have limited descriptions of how the relational component of MI was resolved in their interventions [1]. If MI were to be used as an instrument for psychotherapy, its relational component should be translated in a proper manner to maneuver the treatment to a successful outcome [25]. Though recent studies have attempted to autogenerate words of empathy for a counseling conversation (eg, [26,27]), they provide insufficient description as to how the relational aspect of the conversation should be strategically managed, with much less regard to the specific relational skills.

To address the above problem, this study aimed to incorporate both technical and relational components of MI in a conversational sequence for a brief motivational interview and to investigate its conversational experience on a Web-based text messaging application (*chatbot*). Applying the summons-answer sequence [28], we have built a chatbot that delivers an ordered sequence of MI skills to follow the 4 processes of MI [29] in a conversation with a human user. We focused on the communication between the chatbot and the user, where a smooth interaction is required. The recent mobile chatbot apps that provide therapy (eg, [30-32]) mostly focus on identifying symptoms and providing treatment, leaving the communicative process less attended. Our study has contributed a potential technique and the user conversational experience of translating MI components in a computational manner to broadly inform future Web-based and mobile gadgets that may involve conversational encounters for mental health care.

In this study, a Web-based text messaging application that delivers a brief motivational interview for stress reduction was presented to a group of graduate students for a case study. Graduate students are reported with serious risks of a mental health crisis [33], and a large portion of the graduate student population already ails with mental illnesses such as anxiety and depression. Most of them are reported to have low life satisfaction and even a *tremendous* amount of stress [34,35]; they are 6 times more likely to be exposed to the risk of mental health illnesses than the general population [33]. We designed the chatbot conversation to concern the life of graduate students, instead of addressing all populations [36], to suit a more focused, contextualized conversation.

Research Questions

This study aimed to design a conversational sequence for a brief motivational interview for stress management and to investigate its conversational experience with a group of graduate students on a Web-based text messaging application. The research questions were as follows:

- How do users perceive the sequenced conversation with regard to the MI components?
- What aspects of the conversation can help graduate students cope with stress?
- In what ways can the conversation be improved for better mental health support?

Our study shares much with the computerized MI dialogues in earlier research [13,14] yet differentiates itself in that it is designed for a potential mental health concern and that it uses a sequential organization of MI skills to investigate the user experience of an automated counseling conversation.

Methods

We have illustrated how we translated the theoretical aspects of MI in a Web-based text messaging application called Bonobot. It generates an automated conversational sequence of a brief motivational interview with graduate students for coping with stress.

Technical and Relational Components of Motivational Interviewing in Chatbot

MI entails technical and relational components [1]. The technical component includes counsellor techniques (eg, open questions and reflections) to facilitate change talk, where the client argument for behavior change is formulated [37,38]. The relational component [38], an empathic understanding experienced by clients as the counsellor helps them verbalize change, is also important to the efficacy of MI. In this study, the technical component is translated in a series of MI skills to represent MI counsellor behavior that may evoke change talk. As for the relational component, as Bonobot as a nonhuman agent is inherently incapable of empathy, we tried to achieve such a feat by designing its interaction in the following manner: (1) contextualizing the chatbot responses to the graduate school context [39]; (2) not bombarding questions at the user [29,40]; and (3) applying different combinations of MI skills [41] in the progress of communication.

Though a human MI counsellor would make use of MI-consistent skills spontaneously, a fully natural language conversation is a feat beyond current technology. We have worked around this problem by employing a summons-answer sequence, which can facilitate an exchange of volleys [41] between the summoner and the summoned. Here, the summoning agent asks questions to which the summoned user answers. The agent, in turn, gives feedback. Such an orderliness continues with alternations of volleys between the 2 parties, as in an *abab* formula [28].

Formulating Chatbot Responses

To ensure Bonobot provides responses in appropriate MI skills and communicates them in a proper manner to qualify for both MI components, its responses took the following steps in preparation. First, SHP and JC collected model counsellor statements that may qualify for MI skills from the literature [24,42-48]. Second, they reviewed the statements to gather more generic ones. For example, they removed statements that are narrow-focused (eg, “You’ve been homeless since April...what happened that made your anger reach a breaking point last night?” [45]) and blanked portions of statements to be replaced with fillers from client input (eg, “What was helpful when you feel (client_input_emotion)?” [43]). Finally, to help the agent be more expressive of empathy with regard to the life of a graduate student, SHP and JC modified and added more contextualized statements (eg, “What were your initial goals when you first planned for a graduate degree?”).

A total of 220 prepared statements were later reviewed by certified therapists, CK and SL. They first considered the Motivational Interviewing Skills Code (MISC) [41] to evaluate the responses with regard to MI. As our study primarily concerns the chatbot responses as MI counsellor language, the therapists used the Motivational Interviewing Treatment Integrity (MITI) [49] that refers only to the therapist behavior [38] from the MISC. They coded each statement with the following MITI categories. Examples of the coded responses are provided in Table 1. For predefined (giving information [GI]) responses, see Multimedia Appendix 1.

- *Giving Information(GI).* MI counsellor gives information to educate or provide feedback. As for Bonobot, it provides templated responses to address its role, privacy rules, and the beginning and closing of the session.
- *Questions (Q).* In MI, the counsellor is expected to ask questions that invite elaboration on the problem as well as questions that may evoke change talk. Bonobot uses both types according to the stage of the conversation: *focusing questions (FQ)* and *evoking questions (EQ)*.
- *Reflections (R).* Reflections convey understanding, facilitate exchanges, or further add substantial meaning to what clients say. Bonobot uses simple reflections to acknowledge client remarks and lead the conversation.
- *MI-Adherent Statements (MIA).* MI-adherent statements include any counsellor behavior that is aligned with the MI approach. For Bonobot, we intended affirming statements that may posit client traits in their articulating change.

Sequencing in Chatbot Conversation

Bonobot is built to lead a structured conversation that follows the flow of an MI interaction, namely the 4 processes of MI [29]: Engaging, Focusing, Evoking, and Planning (see Figure 1). Engaging builds a relational foundation with the client. The client’s target behavior is determined in Focusing. In Evoking, change is explored, ideally with the resolution of ambivalence. Planning consolidates client commitment and actions. As Bonobot can only utilize predefined responses, we defined operational aims to reflect the 4 processes within the technical boundaries. In Engaging, Bonobot shares brief introductions with the user and gives instructions to use the chatbot. In Focusing, Bonobot asks the user to detail their problem, possibly having them identify an inner struggle. This leads to Evoking, where Bonobot explores future goals with the user, affirming their own ideas for change. Finally, Bonobot invites the user to ponder the overall session in Planning.

To address the aim of each process, Bonobot uses different combinations of MI skills in each stage (see Figure 1). For the first and last stages, Engaging and Planning, respectively, Bonobot interacts with predefined GI templates to properly manage the beginning and ending of the conversation. In Focusing, FQs are followed with Rs to reveal and reflect on any struggle about the problem. In Evoking, EQs are prompted to encourage change talk and are followed by Rs and MIAs to explore and affirm the idea of change. As advised by the literature [29,40], no more than 2 questions are asked in a row. Rs and MIAs are primarily placed after FQs and EQs as feedback.

Table 1. Examples of Bonobot responses by motivational interviewing (MI) skill.

MI skill	Questions type	Example response
Questions	Focusing questions	In what way does this bother you?
		How would you feel about that?
	Evoking questions	How have you coped with difficult times in the past?
		What were your initial goals when you first planned for a graduate degree?
Reflections	—[a]	It’s tough being a grad student.
		You certainly have a lot on your mind.
MI-adherent statements	—	Sometimes you show a determination that surprises even you.
		It seems like you are a really spirited and strong-willed person in a way.

[a] Not applicable.

Figure 1. The stages and sequence of the Bonobot conversation. EQ: evoking questions; FQ: focusing questions; GI: giving information; MIA: motivational interviewing–adherent statements; R: reflections.

Implementation of the Chatbot Application

Bonobot runs a conversation by generating responses based on keywords. We extended the framework of ELIZA [50], the first chatbot in history, so that Bonobot identifies user keywords but generates responses in the form of an MI skill. We also built 2 modules in the application, Flow Manager and Response Generator, which would execute the sequence and assemble responses.

Preparing Keywords and Responses

A pool of keywords and responses was prepared in a script for Bonobot as shown below:

- *Keywords.* We replaced most keywords in a reproduced ELIZA script [51] with ones obtained from 2 online graduate student communities (r/PhD [52] and r/GradSchool [53]) on Reddit, a social media platform. SHP and SL categorized 1000 posts from each community by open coding for topics based on the title and content [54]. JC reviewed the topics and resolved disagreements with SHP. A word frequency analysis using tf-idf [55] yielded keywords by topic. SHP, JC, and SL went through an iterative process of distributing weights to keywords, so that the graduate school–related ones would be weighted higher. A total of 70 keyword categories were prepared.
- *Responses.* SL programmed responses to be reproduced from the pool of prepared MI statements, triggered by keywords from user input. For each keyword, a designated set of MI skills was allotted. Altogether, with repetition, a total of 209 FQs, 188 EQs, 166 Rs, and 140 MIAs were prepared in the chatbot script. There were 8 GI templates to be used in the beginning and end of the conversation. We also included extra responses to resume the conversation in case no keyword was matched.

Running the Conversation

Bonobot's 2 modules, Flow Manager and Response Generator, were programmed in JavaScript language. Python's Flask 1.0.2 framework was used as the Web application server. The modules work together to run the 4-staged conversation. A series of pilot study sessions informed the final sequencing and turns.

- *Flow Manager.* Flow Manager runs the conversation from one stage to another. At the beginning and end of the conversation, it assigns templated responses to lead the user into and wrap up the conversation. In between, Flow Manager counts the steps in a sequence so that the conversation follows the sequence. If a user does not respond in 10 seconds, it prompts an additional question from Response Generator.
- *Response Generator.* Response Generator identifies keywords and assembles responses (see Figure 2). For instance, suppose a user types in "I don't know if I can graduate." in the Evoking stage. Flow Manager alerts the MI skill to be printed next ("EQ") and Response Generator extracts keywords from the user's input ("I", "know", "if", and "graduate"). It prints the reassembled response ("EQ"; "What changes do you wish to make, if any?") under the highest weighted keyword ("know (5)"). It never repeats the same response twice.

Pilot sessions with 10 graduate students (7 males) aged between 24 and 32 years determined 2 distinct sequences for the Focusing and Evoking stages: (1) to encourage the user to share the problem, an FQ is followed by an R; and (2) to affirm the user's consideration of change, an EQ-R pair is followed by an MIA. In each stage, Bonobot is to repeat the sequence 4 and 6 times, respectively (see Figure 1). This will make up a total of 8 and 18 Bonobot turns in each, with possible extra ones due to the 10-second inactivity rule. Finally, the conversation takes place on a text messaging application in an internet browser (see Figure 3 and Multimedia Appendix 2).

Figure 2. The automated generation of responses via Bonobot modules. EQ: evoking questions; FQ: focusing questions; MIA: motivational interviewing–adherent statements; R: reflections.

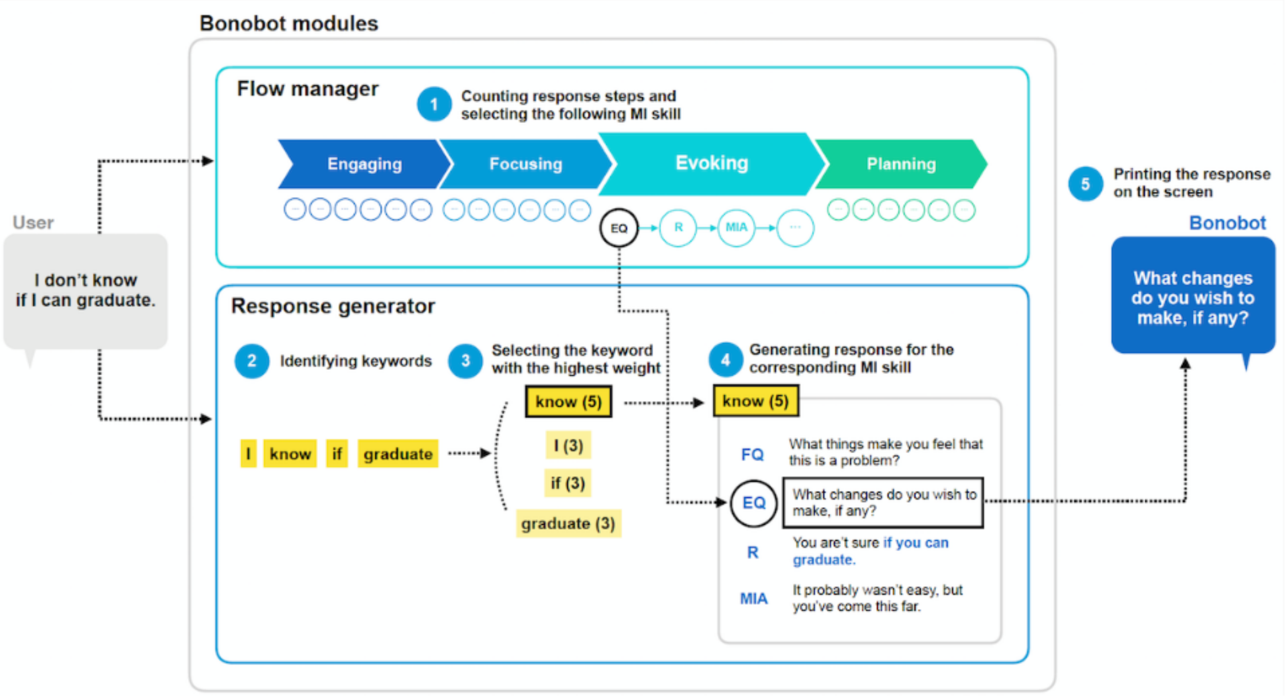
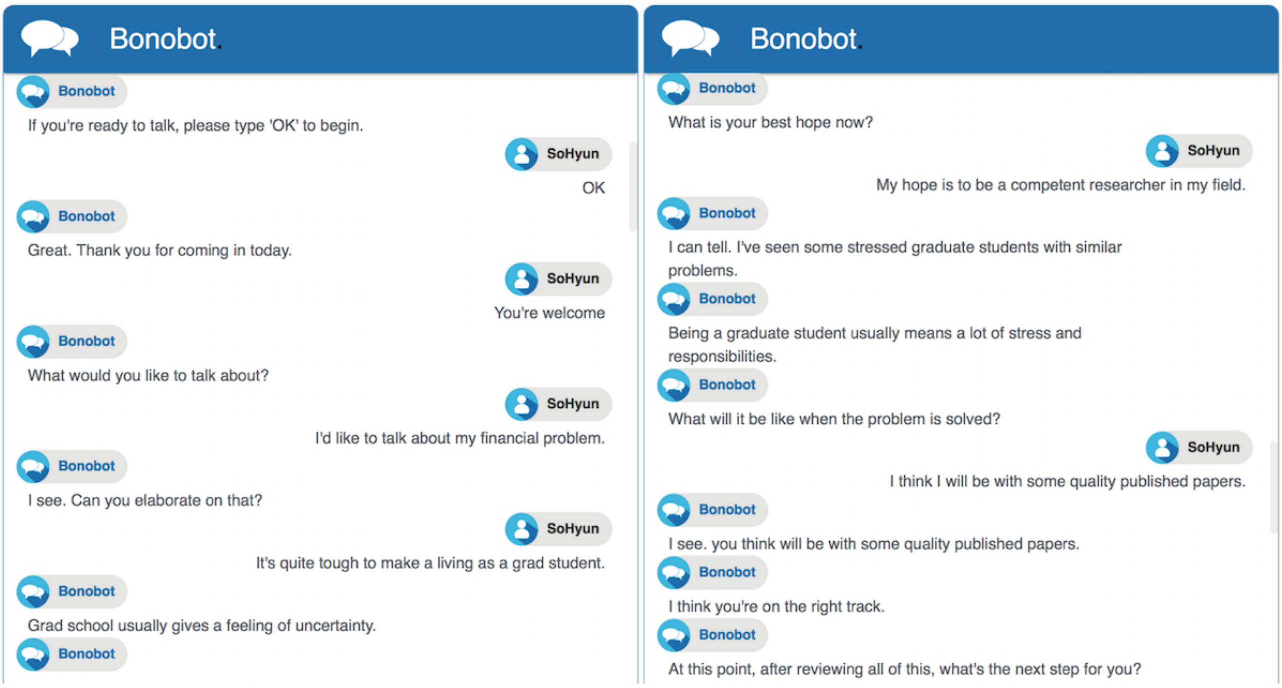


Figure 3. Excerpts from example conversations with Bonobot (For a full example, see Multimedia Appendix 2).


Bonobot

Bonobot
If you're ready to talk, please type 'OK' to begin.

SoHyun
OK

Bonobot
Great. Thank you for coming in today.

SoHyun
You're welcome

Bonobot
What would you like to talk about?

SoHyun
I'd like to talk about my financial problem.

Bonobot
I see. Can you elaborate on that?

SoHyun
It's quite tough to make a living as a grad student.

Bonobot
Grad school usually gives a feeling of uncertainty.

Bonobot

Bonobot

Bonobot
What is your best hope now?

SoHyun
My hope is to be a competent researcher in my field.

Bonobot
I can tell. I've seen some stressed graduate students with similar problems.

Bonobot
Being a graduate student usually means a lot of stress and responsibilities.

Bonobot
What will it be like when the problem is solved?

SoHyun
I think I will be with some quality published papers.

Bonobot
I see. you think will be with some quality published papers.

Bonobot
I think you're on the right track.

Bonobot
At this point, after reviewing all of this, what's the next step for you?

Study Design

A study was designed to investigate (1) the conversational user experience with Bonobot; (2) the impact on their coping with stress; and (3) their needs for better mental health support.

Task and Recruitment

A chatting session with Bonobot was prepared for study participants. An advertisement for volunteers was posted on a Seoul National University online bulletin. A total of 30 full-time graduate students were recruited. The inclusion criteria were that they could (1) communicate with the chatbot in English, (2) share their concerns about school, and (3) participate in an interview about the chatting experience.

Procedure

Participants were invited into a room with a comfortable chair, big table, and laptop computer. A laptop was used instead of the user's mobile phone for consistency and screen convenience. After SHP and JC gave a brief introduction, participants answered a survey of demographic information and the Perceived Stress Scale (PSS-10) [56]. After SHP and JC explained how to use the chatbot, they exited the room for the participants to chat with Bonobot alone. They returned on the participants' notice and conducted semistructured interviews, reviewing the conversation on the laptop screen. The entire process was designed for an hour, and participants received a US $10 beverage coupon as a reward upon completion.

Ethics Approval

Before they gave consent, all participants were informed of the purpose and procedure of the study and that they could resign from it at any point if they felt uncomfortable. The study conformed to the principles of scientific research with human subjects. It was approved by the Seoul National University Institutional Review Board (IRB No. 1708/001-018).

Data Collection and Analysis

In this study, 2 types of data were collected. First, participants' age group, gender, and perceived stress scores were entered on Microsoft Excel by SHP. Perceived stress was measured on a 4-point Likert scale. PSS-10 is one of the most widely used instruments to assess one's perception of stress in the course of the previous month, and higher PSS scores are associated with higher risks to negative health conditions [56]. The collected scores were computed for mean and SD values.

SHP and JC conducted semistructured interviews as student project coordinators and collected the data. They inquired (1) any sequential encounter that went particularly well or oddly and why; (2) whether the conversation had any impact on their perceived stress or not and in what ways; and (3) any suggestions for better mental health support. Detailed questions were asked ad hoc by SHP for an in-depth elaboration on the conversational experience, and notes were taken by JC to record participant-indicated conversational happenings. The interview was designed for 30 min and did not exceed 40 min at most. All interviews were audio-recorded with the consent of the participants and anonymized with numbers.

Interviews were analyzed via a 6-phase process of thematic method by Braun and Clarke [54]. First, all interviews were transcribed verbatim by JC on Microsoft Excel. SHP reviewed and segmented the transcripts by each anonymized participant, using Optimal Workshop's Reframer tool [57]. SHP and JC went through a process of open coding the segments to generate initial codes by tagging each with free-phrased labels. Labels were reviewed and renamed for initial codes. Codes were once again reviewed to search and define themes. As advised by Braun and Clarke [54], themes that merely reiterate the interview questions were avoided but those that can reveal the depth of the data were reviewed and redefined in iteration. Finally, 3 pairs of themes were prepared: Evocative Questions and Unfitting Feedback; Invoking Self-Reflection and Encouragements for Graduate Students; and Need for Informational Support and More Contextualized Feedback. In reporting the results, we rephrased the pairs into 3 overarching themes to answer each research question.

Results

Participant Demographics

Participants were in their 20s (n=20) and 30s (n=10), and a half of them were male (n=15). The average PSS score was 22.5 (SD 5.0), higher than the norm [56]. Conversation topics included the lack of confidence in academic work and research (40%, 12/30), psychological burden of writing theses (17%, 5/30), financial constraints (10%, 3/30), uncertainty about the future (10%, 3/30), work-life balance (7%, 2/30), people skills (7%, 2/30), and other (10%, 3/30). These topics of concern were mostly in agreement (90%, 27/30) with the themes discovered from the content analysis from the Reddit posts.

Qualitative Findings

From the conversation, participants preferred Bonobot's questions to its feedback. EQs were a good means of reflecting on themselves and for some, an instrument for motivational boost. Some of Bonobot's responses related to graduate school were appreciated. All participants favored the idea of using a chatbot for coping with stress, with suggestions for better support. The themes are detailed below.

Evocative Questions and Clichéd Feedback

Participants mostly favored the way Bonobot kept asking them questions. It felt like they were being heard (n=18). In particular, they preferred the EQs in the third stage as they were "something new and interesting" (P2) and "triggered inspiring ideas" (P13). P1 said he liked them as "the questions were profound [...] I had to think deep down and discover the answers inside." Questions such as "What can be some of the good things about making a change?", "What do you wish to be different?", "How would you like things to turn out for you?", and "What could be the next step now?" triggered to think "who you really are and what you really want" (P12) and "what needs to be done to achieve your goals" (P13). P11 said that "it was really the third stage" that "felt quite convenient to draw something out" from him.

However, they did not like questions that made them reiterate their answers. Although Bonobot did not ask the same question twice, participants felt that some questions had them elaborate themselves again. This was a bit annoying to some participants (n=7). In addition, some questions did not feel productive when they were not relevant to the context of their problems and spanned too grand a scheme of things. P4 pointed out an example:

> *Bonobot asked me what I would have chosen to do if I did not pursue a graduate degree. But I've never thought of such an idea—something other than grad school. I'd say that wasn't quite helpful.*

P30 added that "the questions aimed too broad a range that each question could entail a whole lot different story by itself. I think the conversation was too short for that."

In between the questions, Bonobot gave feedback that intended reflections and affirming statements. Most participants (n=21) liked Bonobot saying "such sweet words" (P21). P9 said, "I thought Bonobot used words of empathy really well, you know, even if some felt like templates, they were good." However, it was not quite up to their expectations (n=13). P29 said, "here, Bonobot said the right thing, but it doesn't fit into what I said. I had to doubt whether it really understood me." P2 also said that "I know it tried to encourage me, but sometimes it did at the wrong time, which made me wonder if Bonobot was to encourage me no matter what." P10, P16, and P25 said they anticipated something more than Bonobot simply repeating what they have said. P27 recalled, "It was not bad, but it can be weird… you don't really recite word for word when you talk."

Some feedback, such as "I hear your struggle.", "You certainly have a lot on your mind.", "That's understandable.", and "You're not the only one in this." also felt rather banal. P4 said "Some felt like they were just there because they had to." P21 and P24 found it odd that Bonobot repeated similar expressions in the conversation. In addition, they were just "too nice" (P2). P1 added that "you know, if you were to talk with a human being, you wouldn't really say the nicest things throughout." P4 still "appreciated the niceness" as he rarely has a chance for those words. However, for P17, P20, P24, and P28, words of empathy only echoed what they could expect from anybody around them.

Self-Reflection and Potential Consolation

Most found the Evoking stage an opportunity to reflect on themselves (n=19). P4 said "I think it was a time to reflect on me and my situation. [...] I liked I had the chance to rediscover myself with my own words by chatting with Bonobot." For P22, it was "a Socratic method": "In the end, you answer for yourself. Bonobot asks me questions, and by answering them, I get a better understanding of myself." P23 also said it was like "a catalyst" that kept nudging her to think about herself and her life. This self-reflection spurred a sense of motivation (n=11) that was "sort of buried in" (P27). P30 said she could gain a motivational boost:

> *You know, I'm always like, "what am I to do now?," "this is too hard," and "I can't do it." Now, I have this question inside, "so how do I want this to be resolved?" This can move me forward. I feel like I need to do something about it.*

The progression toward Evoking, however, was not everyone's preference (n=6). P17 said, "Bonobot clearly had an idea about what it wanted to hear from me—something positive—and it wanted me to say it, which made me feel like Bonobot had the lead over the conversation, not me." P18 also disliked the idea of "having a conversation with a purpose." P20, in particular, had trouble facing the idea of change:

> *I guess I am not exactly sure. I know I need to change, and I know what I need to do to make that change happen. But that's causing the stress! But Bonobot's questions felt like it was trying to remind me of that, instead of letting me vent.*

Participants appreciated Bonobot taking the role of a nonjudgmental listener (n=11). P3 thought she made "a virtual friend who listens to [her] and tries to understand [her]." It was essentially a private conversation where they could talk about things that they cannot usually open up to their family or friends (n=8). P19 said he could feel more relaxed talking to Bonobot "for [he] did not have to worry about what Bonobot would think of [him]." P24 said he shared the same subject that he did with a colleague, which ended up in an argument. He felt better talking with Bonobot "for [Bonobot] does not have any interests that may conflict with mine."

They preferred Bonobot's words of empathy that concerned the life in graduate school (n=13). P3 said that Bonobot seemed to know "what it is to be a graduate student." P6 felt touched when it said, "Don't let it discourage you," to his disconsolation with his progress. When Bonobot asked P7 about the past achievements to which he had none, it replied "That's okay," with which he felt so moved and said, "it understood me like a human being!" P23, in particular, was pleased with Bonobot saying "A lot of graduate students suffer from variations of the same problem.", when she confessed trouble with her advisor. Similarly, P15's favorite was "That's exactly how many students feel during their graduate program." P9, P19, and P30 also indicated the following as the best responses: "Grad school usually gives a feeling of uncertainty." and "It's tough being a grad student."

Need for Information and Contextualized Feedback

Talking about making a change brought about a need for data and information (n=9). P7 said "I need more information I guess, about the problem I talked about." P8 said the conversation would have helped "if the chatbot offered tips for writing theses on the Internet." P26 suggested the following:

> *a chatbot can deliver news articles or life tips. You know, say I have a sleeping problem. It can give me suggestions, such as music recommendations, health information, or other tips found online.*

Some participants also indicated a need for making plans for change (n=5). P13 recommended that Bonobot ask more detailed, branching questions such as "How much financial aid do you get?" or "What are the current career options?". P5 said "I would appreciate it much more if it organized for me a listed reminder of the things that were brought up in the conversation." P14 and P22 also suggested that discussing specific action items would be helpful. P26 said that planning an agenda with Bonobot would potentially inspire a sense of partnership.

Most participants said they would prefer having an in-depth conversation, digging deeper a single subject instead of a range of things (n=17). P9 thought the conversation tried to cover too much:

> *if I were to talk about the same subject with my friend, we would talk about it for a long time. But in this conversation, it feels like I need to share just enough and then move on right away to "think about the next, and the next." I wish it would rather ask me more details about my problem.*

Other participants also agreed that they would want a chatbot to ask them more specific questions about their life, "as if [they were] talking to a human being" (P20). Addressing such an elaborated context of their problems would signal "a continued relationship" (P5) with the chatbot (n=7).

In addition, participants wanted more emotional responses that are appropriately contextualized to their input (n=13). P12 said, “This chatbot can say some sweet words, but I would prefer more emotional expressions like, ‘I can’t believe that happened to you!’ or ‘That must have been very hard on you,’ things like that.” P27 put it this way: “You know, I’d like words that are more for me and me only, not like the mundane ones that anybody can say to everyone else.” For P25, more personalized responses would have helped her feel more empathized:

> *What if it said something more concrete, like, “You must have had a hard time communicating with your advisor all this while,” instead of just a simple expression of empathy? Then I would think that it really understands my feelings.*

Discussion

Principal Findings

This case study has explored conversational experience of an automated sequence of MI skills for stress management on a Web-based text messaging application, with an aim to integrate both technical and relational components of MI. Participants found the sequenced conversation quite natural and easy to grasp and follow. The evocative questions inspired discussions about change for most, encouraging self-reflection. Some agent-generated feedback that reflected their life in school was appreciated. Participants suggested informational support for the chatbot to equip a better preparation tool for helping change. A need for an in-depth conversation was also indicated, with more contextualized feedback for better engagement.

Our study highlights the possibility of applying a sequential approach to constructing an automated motivational interview with MI skills to integrate both technical and relational components. Previous research has remained rather vague in explicating the relational component or has excluded it in its entirety owing to technological issues (ie, [1,12]). It is important that the MI counsellor uses interactional skills strategically to convey an empathic understanding [15]. However, not much has been studied on how one can translate such a technique in a computational manner. We designed a structured conversation with a sequence of MI skills, with an effort to incorporate both components of MI. Using the summons-answer sequence [28], we placed questions sparingly and not consecutively [29] and assigned reflections and MI-adherent statements in-between to form the basis for an empathic understanding [39,41]. The result was the FQ-R and EQ-R-MIA sequences in the second and third stages, respectively, with GI templates at the beginning and at the end.

The sequence demonstrated a reasonable, though not optimal, MI interaction. The questions conveniently had the user talk about their problems, and the conversation encouraged a chance for self-reflection for most and inspired an idea of change for some participants. However, agent-generated reflections and MI-adherent statements received mixed opinions. Owing to its technical constraints as a chatbot counsellor, some of Bonobot’s responses were ill-assembled and could not correctly reflect or affirm on user volleys. Although participants liked the moral support from the chatbot’s feedback, we caution that it may risk sustaining the status quo instead of change [58]. An accurate empathy requires a profound contextual understanding, which is hardly achieved by nonhuman agents. Further research is needed to generate chatbot responses that are appropriately tailored as well as MI-consistent to avoid naively echoing client remarks in reflections and simply abstracting them in questions.

Participants regarded evocative questions as a constructive means to revisit their source of stress, leading to the idea of change. In the interview, participants who were able to ponder change were willing to share their immediate plans to cope. However, for some, the distaste and even resistance to problem-solving actions was also observed. We find both types of reactions to be in alignment with the literature [38], and highlight the potential influence of change talk on stress coping behavior. The Evoking stage could encourage self-reflection, potentially playing a part in coping with stress. It prompted participants to think why the problem is stressful and how they want it to be resolved. In terms of Lazarus and Folkman’s transactional model [59], this process is likely a cognitive *reappraisal* of the stressful condition. Such a positive reinterpretation is not only a means to reduce emotional distress but also a form of active, problem-focused coping [60]. This finding leads to a future research agenda to collect concrete evidence of change talk in chatbot-client conversations and measure its empirical effect on stress reduction as a coping intervention.

What participants requested for better mental health support suggests the potential of a chatbot counsellor, as well as milestones to be achieved in technology. Our study has made a meaningful step forward to address both components of MI in a chatbot app, supporting a real-time speaker exchange while utilizing different MI skills. As for the technical component, the results show that MI techniques were in action. Participants needed more informational support as they revisited their problem and began planning to cope. The chatbot app needs to be equipped with problem-related information in the future. In terms of the relational component, participants agreed on Bonobot’s caring attitude, a ground hypothesis for a client-centered approach [61]. However, the need for better contextualized feedback demands much advance in technology to generate intelligent, context-aware chatbot responses that can contribute to client change talk. Although the chatbot in our study could not fully exhaust the MI counsellorship, future research can make use of advanced algorithms such as artificial neural networks to generate more sophisticated chatbot responses that better reflect the theory and principles.

Our findings lead us to suggest that, if properly and carefully designed, a chatbot may conveniently serve the purpose of MI as an *interactional* practice in health [62,63]. As a real-time messaging application, chatbots can help tend communication for a therapeutic encounter between a counsellor and client. The recent chatbot apps that provide therapy (eg, [30-32]) mainly serve the role of delivering various treatment programs via a conversation. Our study stresses on the conversation with the chatbot itself as the potential medium to render a motivational interview, for mental health concerns in particular. As we face an unprecedently technology-intensive era, we foresee a number of conversational agents to appear in the communicative process of providing care (eg, [64-68]). To properly manage such an interaction, we believe a well-designed conversational sequence is necessary. We suggest having a close collaboration with human-computer interaction in studying the internet and mobile health care technologies [69] to address many complex problems that may arise with users verbally engaging in the technologically adapted treatment.

Limitations and Directions for Future Research

Our work is bound with limitations. First, Bonobot utilized 1 possible sequence of MI skills. A gamut of sequences can be tried for further investigation. In addition, there are a number of factors that may have contributed to the potential impact of the conversation, for example, the number of chatbot responses and order of skills to name just a few. More sophisticated designs of MI conversations can be explored in the future.

Moreover, Bonobot could only support a single session. We believe evocative questions made a positive impression for many to begin with. However, it could not be fully explored, along with the ambivalence associated with it, owing to technical and circumstantial constraints. A continued, fully fledged MI session should follow to investigate the longitudinal effect of change talk as a coping instrument.

As a case study, our research presents a small participant sample, and hence, the scope and range of stressors of graduate students could not be exhausted. A stressor can be a complex, multi-layered problem, spanning various aspects of one's life. Setting the scope of a stressful state context would be an important challenge to be resolved in the future, before a larger field study can be conducted.

This study also lacks a quantifiable measure in assessing the chatbot conversation. Although there are great inventories for evaluating client-counsellor or doctor-patient relationships (eg, [70,71]), few exist for human-agent interaction in health care encounters. Future research, possibly in human-robot interaction, could address this need for a toolkit to assess client-chatbot conversations.

Finally, other types of multimedia, in the form of animated characters, robots, avatars, or other embodiments, were suggested by a few participants for more affability and sociability (P5, P20, and P23). Our research only allowed text-based communication, even without textual emojis, to control for any effect other than from the sequence. It would be interesting to explore the effectiveness of multimedia resources or embodiment features on the relational component of MI.

Conclusions

This case study has designed a conversational sequence of MI skills for an automated motivational interview on stress management and presented qualitative feedback from 30 participants on their conversational experience. Our findings revealed user preference for evocative questions but less inclination for agent-generated reflective and affirming feedback. Sequencing questions and MI-adherent statements can lead a conversation for coping with stress, possibly encouraging self-reflection. Participants demanded informational support as well as more contextualized words of empathy. Our study contributes to technological adaptations of MI and informs the design of future conversational agents in mental health care.

Acknowledgments

This work was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government (No.2017-0-00693), and the National Research Foundation of Korea (NRF) grant funded by the Korean National Police Agency and the Ministry of Science and ICT for police research and development (NRF-2018M3E2A1081492).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bonobot responses by motivational interviewing skill.

[PDF File (Adobe PDF File), 68KB - jmir_v21i4e12231_app1.pdf]

Multimedia Appendix 2

Example conversation with Bonobot.

[PDF File (Adobe PDF File), 268KB - jmir_v21i4e12231_app2.pdf]

References

1. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
2. Blanson Henkemans OA, van der Boog PJ, Lindenberg J, van der Mast CA, Neerincx MA, Zwetsloot-Schonk BJ. An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. Technol Health Care 2009;17(3):253-267. [doi: 10.3233/THC-2009-0545] [Medline: 19641261]

3. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
4. Gerbert B, Berg-Smith S, Mancuso M, Caspers N, McPhee S, Null D, et al. Using innovative Video Doctor technology in primary care to deliver brief smoking and alcohol intervention. Health Promot Pract 2003 Jul;4(3):249-261. [doi: 10.1177/1524839903004003009] [Medline: 14610995]
5. Leung SF, Ma J, Russell J. Enhancing motivation to change in eating disorders with an online self-help program. Int J Ment Health Nurs 2013 Aug;22(4):329-339. [doi: 10.1111/j.1447-0349.2012.00870.x] [Medline: 22882349]
6. Ondersma SJ, Svikis DS, Schuster CR. Computer-based brief intervention: a randomized trial with postpartum women. Am J Prev Med 2007 Mar;32(3):231-238 [FREE Full text] [doi: 10.1016/j.amepre.2006.11.003] [Medline: 17236741]
7. Webber KH, Tate DF, Michael BJ. A randomized comparison of two motivationally enhanced Internet behavioral weight loss programs. Behav Res Ther 2008 Sep;46(9):1090-1095. [doi: 10.1016/j.brat.2008.06.008] [Medline: 18675402]
8. Jackson RA, Stotland NE, Caughey AB, Gerbert B. Improving diet and exercise in pregnancy with Video Doctor counseling: a randomized trial. Patient Educ Couns 2011 May;83(2):203-209. [doi: 10.1016/j.pec.2010.05.019] [Medline: 21459255]
9. Tsoh JY, Kohn MA, Gerbert B. Promoting smoking cessation in pregnancy with Video Doctor plus Provider Cueing: a randomized trial. Acta Obstet Gynecol Scand 2010 Apr;89(4):515-523 [FREE Full text] [doi: 10.3109/00016341003678419] [Medline: 20196678]
10. Humphreys J, Tsoh JY, Kohn MA, Gerbert B. Increasing discussions of intimate partner violence in prenatal care using Video Doctor plus Provider Cueing: a randomized, controlled trial. Womens Health Issues 2011 Mar;21(2):136-144 [FREE Full text] [doi: 10.1016/j.whi.2010.09.006] [Medline: 21185737]
11. Kanaoka T, Mutlu B. Designing a motivational agent for behavior change in physical activity. In: CHI EA '15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2015 Presented at: The 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems; April 18-23, 2015; New York, NY p. 1445-1450. [doi: 10.1145/2702613.2732924]
12. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a motivational interview delivered by a robot: qualitative study. J Med Internet Res 2018 May 03;20(5):e116 [FREE Full text] [doi: 10.2196/jmir.7737] [Medline: 29724701]
13. Schulman D, Bickmore T, Sidner C. An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In: AAAI Spring Symposium, Technical Report. 2011 Presented at: AAAI Spring Symposium; March 21-23, 2011; Palo Alto, CA.
14. Imperato C, Hempstead M, Schulman D, Kopy M, Tran HS, Yinusa-Nyahkoon L, et al. Incorporating motivational interviewing into virtual contraceptive counseling. 2012 Presented at: The 140th APHA Annual Meeting and Exposition; 2012; San Francisco, CA.
15. Rollnick S, Miller WR. What is motivational interviewing? Behav Cognit Psychother 2009 Jun 16;23(04):325-334. [doi: 10.1017/S135246580001643X]
16. Westra HA, Aviram A, Doell FK. Extending motivational interviewing to the treatment of major mental health problems: current directions and evidence. Can J Psychiatry 2011 Nov;56(11):643-650. [doi: 10.1177/070674371105601102] [Medline: 22114918]
17. Arkowitz H, Westra HA. Introduction to the special series on motivational interviewing and psychotherapy. J Clin Psychol 2009 Nov;65(11):1149-1155. [doi: 10.1002/jclp.20640] [Medline: 19785011]
18. Buckner JD, Ledley DR, Heimberg RG, Schmidt NB. Treating comorbid social anxiety and alcohol use disorders: combining motivation enhancement therapy with cognitive-behavioral therapy. Clin Case Stud 2008 Jun;7(3):208-223 [FREE Full text] [doi: 10.1177/1534650107306877] [Medline: 20414473]
19. Arkowitz H, Westra HA. Integrating motivational interviewing and cognitive behavioral therapy in the treatment of depression and anxiety. J Cogn Psychother 2004 Oct 01;18(4):337-350. [doi: 10.1891/jcop.18.4.337.63998]
20. Westra HA. Managing resistance in cognitive behavioural therapy: the application of motivational interviewing in mixed anxiety and depression. Cogn Behav Ther 2004;33(4):161-175. [doi: 10.1080/16506070410026426] [Medline: 15625790]
21. Westra HA, Dozois DJ. Preparing clients for cognitive behavioral therapy: a randomized pilot study of motivational interviewing for anxiety. Cogn Ther Res 2006 Aug 11;30(4):481-498. [doi: 10.1007/s10608-006-9016-y]
22. Burke BL. What can motivational interviewing do for you? Cogn Behav Pract 2011 Feb;18(1):74-81. [doi: 10.1016/j.cbpra.2009.08.004]
23. Fried RR, Irwin JD. Calmly coping: a motivational interviewing via CoActive Life Coaching (MI-VIA-CALC) pilot intervention for university students with perceived levels of high stress. International Journal of Evidence Based Coaching and Mentoring 2016;14(1):16-33 [FREE Full text]
24. Miller W, Rollnick S. Motivational Interviewing: Preparing People for Change. New York, NY: Guilford Publications; 2002.
25. Westra HA. Motivational Interviewing in the Treatment of Anxiety. New York, NY: Guilford Press; 2012.

26. Oh K, Lee D, Ko B, Choi H. A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. 2017 Presented at: The 18th IEEE International Conference on Mobile Data Management (MDM); May 29-June 1, 2017; Daejeon, Korea p. 371-375. [doi: 10.1109/MDM.2017.64]
27. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
28. Schegloff EA. Sequencing in conversational openings. American Anthropologist 1968 Dec;70(6):1075-1095 [FREE Full text] [doi: 10.1525/aa.1968.70.6.02a00030]
29. Miller WR, Rollnick S. Motivational interviewing: Helping People Change. New York, NY: Guilford Press; 2013.
30. 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]
31. HelloJoy. 2018. URL: http://www.hellojoy.ai/ [accessed 2018-08-31] [WebCite Cache ID 724mDyWB7]
32. 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 2018:6-(11) [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
33. Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. Evidence for a mental health crisis in graduate education. Nat Biotechnol 2018 Mar 06;36(3):282-284. [doi: 10.1038/nbt.4089] [Medline: 29509732]
34. Smith E, Brooks Z. National Association of Graduate-Professional Students. 2015. Graduate student mental health 2015 URL: http://nagps.org/wordpress/wp-content/uploads/2015/06/NAGPS_Institute_mental_health_survey_report_2015.pdf [accessed 2018-08-01] [WebCite Cache ID 71LND2gfh]
35. University of California Office of the President. 2017. The University of California Graduate Student Well-Being Survey Report URL: https://www.ucop.edu/institutional-research-academic-planning/_files/graduate_well_being_survey_report.pdf [accessed 2018-08-01] [WebCite Cache ID 71LNFLnIa]
36. Tambuyzer E, Pieters G, Van Audenhove C. Patient involvement in mental health care: one size does not fit all. Health Expect 2014 Feb;17(1):138-150 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00743.x] [Medline: 22070468]
37. Magill M, Gaume J, Apodaca TR, Walthers J, Mastroleo NR, Borsari B, et al. The technical hypothesis of motivational interviewing: a meta-analysis of MI's key causal model. J Consult Clin Psychol 2014 Dec;82(6):973-983 [FREE Full text] [doi: 10.1037/a0036833] [Medline: 24841862]
38. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
39. Moyers TB, Miller WR. Is low therapist empathy toxic? Psychol Addict Behav 2013 Sep;27(3):878-884 [FREE Full text] [doi: 10.1037/a0030274] [Medline: 23025709]
40. Arkowitz H, Westra H, Miller W, Rollnick S. Motivational Interviewing in the Treatment of Psychological Problems. New York, NY: Guilford Press; 2007.
41. Houck J. The University of New Mexico. 2008. Manual for the Motivational Interviewing Skill Code (MISC) URL: https://casaa.unm.edu/download/misc.pdf [accessed 2019-01-28] [WebCite Cache ID 75oKooaM5]
42. Miller WR, Rollnick S. Motivational Interviewing: Preparing People to Change Addictive Behavior. New York, NY: Guildford Press; 1991.
43. Corcoran J. Motivational Interviewing: A Workbook for Social Workers. New York, NY: Oxford University Press; 2016.
44. LeCompte A. Creating Harmonious Relationships: A Practical Guide to the Power of True Empathy. Portsmouth, NH: Atlantic Books; 2000.
45. Douaihy A, Kelly T, Gold M, editors. Motivational Interviewing: A Guide for Medical Trainees. New York, NY: Oxford University Press; 2014.
46. Chen M, Giblin N. Individual Counseling and Therapy: Skills and Techniques. New York, NY: Routledge; 2017.
47. Coleman H, Yeh C, editors. Handbook of School Counseling (Counseling and Counselor Education). New York, NY: Routledge; 2008.
48. Macdonald A. Solution-Focused Therapy: Theory, Research & Practice. Los Angeles, CA: SAGE Publications, Ltd; 2011.
49. Moyers TB, Martin T, Manuel JK, Hendrickson SM, Miller WR. Assessing competence in the use of motivational interviewing. J Subst Abuse Treat 2005 Jan;28(1):19-26. [doi: 10.1016/j.jsat.2004.11.001] [Medline: 15723728]
50. Weizenbaum J. ELIZA--A computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
51. Landsteiner N. Elizabot. 2005. URL: https://www.masswerk.at/elizabot/ [accessed 2018-08-01] [WebCite Cache ID 71LORSHiP]
52. Reddit. 2018. r/PhD URL: http://www.reddit.com/r/PhD [accessed 2018-09-03] [WebCite Cache ID 729tBW6el]
53. Reddit. 2018. r/GradSchool URL: http://www.reddit.com/r/GradSchool [accessed 2018-09-03] [WebCite Cache ID 729tExr7f]
54. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

55. Ramos J. Using TF-IDF to determine word relevance in document queries. 2003 Presented at: The 1st Instructional Conference on Machine Learning; 2003; Piscataway, NJ p. 133-142.
56. Cohen S. Mind Garden. 1994. Perceived stress scale URL: http://mindgarden.com/documents/PerceivedStressScale.pdf [accessed 2018-08-01] [WebCite Cache ID 71LPDnXRj]
57. Optimal Workshop. 2018. Reframer URL: https://www.optimalworkshop.com/reframer [accessed 2018-09-06] [WebCite Cache ID 72ETtzsQp]
58. Moyers TB, Rowell LN, Manuel JK, Ernst D, Houck JM. The Motivational Interviewing Treatment Integrity code (MITI 4): rationale, preliminary reliability and validity. J Subst Abuse Treat 2016 Dec;65:36-42 [FREE Full text] [doi: 10.1016/j.jsat.2016.01.001] [Medline: 26874558]
59. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York, NY: Springer Publishing Company; 1984.
60. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989;56(2):267-283. [doi: 10.1037/0022-3514.56.2.267]
61. Rogers CR. The necessary and sufficient conditions of therapeutic personality change. J Consult Psychol 1957 Apr;21(2):95-103. [Medline: 13416422]
62. Maynard DW, Heritage J. Conversation analysis, doctor-patient interaction and medical communication. Med Educ 2005 Apr;39(4):428-435. [doi: 10.1111/j.1365-2929.2005.02111.x] [Medline: 15813766]
63. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
64. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, et al. SimSensei kiosk: a virtual human interviewer for healthcare decision support. In: AAMAS '14.: International Foundation for Autonomous Agents and Multiagent Systems; 2014 Presented at: International Conference on Autonomous Agents and Multi-agent Systems; May 05-09, 2014; Richland, SC.
65. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
66. Utami D, Bickmore T, Kruger L. A robotic couples counselor for promoting positive communication. : IEEE; 2017 Presented at: The 26th IEEE The International Symposium on Robot and Human Interactive Communication (RO-MAN); 28 August-1 September; Lisbon, Portugal p. 248-255. [doi: 10.1109/ROMAN.2017.8172310]
67. 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]
68. Stieger M, Nißen M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: Study protocol of a randomized, wait-list controlled trial. BMC Psychol 2018 Sep 04;6(1):43 [FREE Full text] [doi: 10.1186/s40359-018-0257-9] [Medline: 30180880]
69. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans on Computer-Human Interaction (TOCHI) 2005;12(2):293-327.
70. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Counsel Psychol 1989;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
71. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract 2004 Dec;21(6):699-705. [doi: 10.1093/fampra/cmh621] [Medline: 15528286]

Abbreviations

EQ: evoking questions
FQ: focusing questions
GI: giving information
MI: motivational interviewing
MIA: MI-adherent statements
MISC: Motivational Interviewing Skills Code
MITI: Motivational Interviewing Treatment Integrity
PSS: Perceived Stress Scale
Q: questions
R: reflections

Edited by G Eysenbach; submitted 19.09.18; peer-reviewed by B Borsari, L Lechner; comments to author 06.12.18; revised version received 30.01.19; accepted 17.02.19; published 16.04.19.

Please cite as:
Park S, Choi J, Lee S, Oh C, Kim C, La S, Lee J, Suh B
Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study
J Med Internet Res 2019;21(4):e12231
URL: https://www.jmir.org/2019/4/e12231/
doi:10.2196/12231
PMID:30990463

Original Paper

The Evidence Base for an Ideal Care Pathway for Frail Multimorbid Elderly: Combined Scoping and Systematic Intervention Review

Gro Berntsen[1,2], MD, PhD; Frode Strisland[3], Dr Ing; Kristian Malm-Nicolaisen[1], MSc; Berglind Smaradottir[4,5], PhD; Rune Fensli[4], PhD; Mette Røhne[3], PhD

[1]Norwegian Center for E-health Research, University Hospital of North Norway, Tromsø, Norway
[2]Department of Primary Care, Institute of Community Medicine, UiT-The Arctic University of Norway, Tromsø, Norway
[3]SINTEF Digital, Oslo, Norway
[4]Centre for eHealth, University of Agder, Grimstad, Norway
[5]Research Department, Sørlandet Hospital, Kristiansand, Norway

Corresponding Author:
Gro Berntsen, MD, PhD
Norwegian Center for E-health Research
University Hospital of North Norway
PB 35
9038 Tromsø
Tromsø, 9038
Norway
Phone: 47 90518895
Email: gro.rosvold.berntsen@ehealthresearch.no

Abstract

Background: There is a call for bold and innovative action to transform the current care systems to meet the needs of an increasing population of frail multimorbid elderly. International health organizations propose complex transformations toward digitally supported (1) Person-centered, (2) Integrated, and (3) Proactive care (Digi-PIP care). However, uncertainty regarding both the design and effects of such care transformations remain. Previous reviews have found favorable but unstable impacts of each key element, but the maturity and synergies of the combination of elements are unexplored.

Objective: This study aimed to describe how the literature on whole system complex transformations directed at frail multimorbid elderly reflects (1) operationalization of intervention, (2) maturity, (3) evaluation methodology, and (4) effect on outcomes.

Methods: We performed a systematic health service and electronic health literature review of care transformations targeting frail multimorbid elderly. Papers including (1) Person-centered, integrated, and proactive (PIP) care; (2) at least 1 digital support element; and (3) an effect evaluation of patient health and/ or cost outcomes were eligible. We used a previously published ideal for the quality of care to structure descriptions of each intervention. In a secondary deductive-inductive analysis, we collated the descriptions to create an outline of the generic elements of a Digi-PIP care model. The authors then reviewed each intervention regarding the presence of critical elements, study design quality, and intervention effects.

Results: Out of 927 potentially eligible papers, 10 papers fulfilled the inclusion criteria. All interventions idealized Person-centered care, but only one intervention made *what mattered to the person* visible in the care plan. Care coordinators responsible for a whole-person care plan, shared electronically in some instances, was the primary integrated care strategy. Digitally supported risk stratification and management were the main proactive strategies. No intervention included workflow optimization, monitoring of care delivery, or patient-reported outcomes. All interventions had gaps in the chain of care that threatened desired outcomes. After evaluation of study quality, 4 studies remained. They included outcome analyses on patient satisfaction, quality of life, function, disease process quality, health care utilization, mortality, and staff burnout. Only 2 of 24 analyses showed significant effects.

Conclusions: Despite a strong common-sense belief that the Digi-PIP ingredients are key to sustainable care in the face of the silver tsunami, research has failed to produce evidence for this. We found that interventions reflect a reductionist paradigm, which forces care workers into standardized narrowly focused interventions for complex problems. There is a paucity of studies that meet complex needs with digitally supported flexible and adaptive teamwork. We predict that consistent results from care transformations for frail multimorbid elderly hinges on an individual care pathway, which reflects a synergetic PIP approach enabled by digital support.

(J Med Internet Res 2019;21(4):e12517) doi:10.2196/12517

KEYWORDS

systematic review; patient-centered care; delivery of health care, integrated; secondary prevention; risk management

Introduction

The Perfect Storm

The combination of increased longevity, more sensitive diagnostics, and improved treatment contribute to the increasing prevalence of multimorbidity [1-3]. The link between multimorbidity and higher health care spending is well documented, and, most interestingly, the top 10% spenders, who account for two-thirds of all care spending, are dominated by a group of patients with multimorbidity and complex long-term needs [4-6]. Persons with complex long-term needs require coordinated and seamless care from many different providers. However, although other sectors have adopted digital solutions to *glue* fragmented service processes together and enhance both the efficiency and quality of their service, health care lags behind [7]. Following Conway's law, current information and communication technology infrastructures in health tend to mirror and solidify the silo structures of the organizations they serve [8]. The siloed structure adds to the organizational fragmentation and jeopardizes the information flow [9]. Consequently, the cost and quality of care for persons with complex long-term needs suffer from disruptions, gaps, and duplication of care. Rising expenses, improved single disease treatments, increased proportion of patients with complex long-term needs, and slow adoption of supportive technology are creating a perfect storm threatening the sustainability of our health care systems [10].

Emerging Responses to New Demands

There is a call for bold and innovative action to reform the current analog, episodic, single disease, profession-centric, and reactive care system to meet the new needs of the population. Health care organizations and researchers in the United States, the World Health Organization, and the European Union are developing roadmaps to deal with the silver tsunami [11-14]. These agree on certain central tenets: we need to improve person-centered care and patient engagement, both because it is the right thing to do and because patient involvement and self-management hold the promise of better and more cost-efficient care [15]. Care fragmentation should be met with integration and seamless digital care plans [16]. Proactive and preventive practices will decrease the need for costly care in both human and economic terms and improve outcomes [17,18]. Last but not least, digital tools are essential to leverage new care models and enable scalability [19-21].

However, the agreement and shared understanding of what *Digitally supported PIP care* is and how it is implemented stop there. Although this strategy seems reasonable, its evidence base is unclear. There are some inspiring success stories around digitally supported large-scale system transformations, notably from the United States with Kaiser Permanente [22,23], South Central Foundation [24], and Veterans' health administration (VA) [25]. However, these transformations and effects have been tricky to reproduce [26-28], and there is little consensus on the critical elements for success [16,29-32]. Each of the *PIP care* elements has been subject to systematic reviews with mainly encouraging results [16,18,19,31,33-36]. However, each component is often studied on its own, so that the maturity of each element in comparison with other interventions in the same vein and the synergies between them have not been subject to academic study. There is much literature on promising digital solutions, but the large-scale adoption of these tools is slow [37], and there is disillusionment with large-scale electronic health (eHealth) impact [38]. For simplicity in this paper, we will call this triad Digi-PIP care, an abbreviation for *Digi*tally supported *P*erson-centered, *I*ntegrated, and *P*roactive care.

Research Questions

There is an urgent need to determine whether health services that take a synergistic Digi-PIP care approach to meet the silver tsunami have a documented effect on the triple aims of population health, patient experience, and cost-effectiveness. As care systems vary significantly in the extent they support and operationalize Digi-PIP-care, the review of the documentation is only meaningful if we can get a grasp on the degree of fidelity and maturity of any given implementation.

Our research questions are as follows:

1. How are Digi-PIP care interventions operationalized?
2. How can we capture the maturity of a Digi-PIP intervention?
3. What is the Digi-PIP study methodological quality, and which effects are reported?
4. Does intervention maturity matter for effects?

Hypotheses and Approach

Defining Digitally Supported Person-Centered, Integrated, and Proactive Care

Understanding whether each of the Digi-PIP elements are present or not, and to what degree they are present, is essential to be able to evaluate their effectiveness. In a *drug* model analogy, this would be equivalent to understanding not only if the active drug ingredient is present but also what dosage and administration route is most effective and has fewest side effects. Like in the drug case, we need to grasp the fundamental mechanisms that are responsible for results in the care of patients with complex long-term needs.

However, the individual PIP elements are all poorly defined in the literature. "The term 'person-centered care' is used to refer to many different principles and activities, and there is no single agreed definition of the concept.” [39] Likewise, the terms *integrated care* and *proactive care* suffer from the same conceptual ambiguities. Moreover, other terms are synonymous or overlapping with the PIP terms chosen here, such as client-centered care, continuity of care, and anticipatory care [40].

Instead of trying to harmonize definitions from the literature, or inviting professionals to discuss their way to consensus, we have chosen the patient perspective to be our guiding principle. In a previous paper, we show that to patients with complex long-term needs, the essence of care quality was that the care supported their long-term goals, linked to the question: *What matters to you?* [41]. Furthermore, the concept of the individual Patient Pathways as a cycle of Goal setting > Care planning > Care delivery > Care evaluation > Goal adjustment, and so on, made sense to multimorbid patients. All care pathways can be described in these terms, although the fragmented nature of care organization can make this simple pattern difficult to recognize, as each silo will tend to elicit separate parallel individual patient pathways. In Figure 1, the PIP elements are embedded in the 4 stages of the framework.

Below, we describe each of the active elements in PIP care regarding their key characteristics, their care component, and relevant digital support. See also Figure 1, inspired by Coulter [42].

Person-Centered Care

Defining characteristics: To ensure that care decisions are made in alignment with the person’s answer to the question: *What matters to me?* and that what matters is framed within the scope of realistic, relevant, and safe practices. Common goals have a coordinating effect. The care process is successful when it meets the personal goals [43].

Care component: The pathway starts with a sensitive and empathic exploration of *what matters*. A complex care process without a common goal will quickly become a quagmire of competing and poorly coordinated subprocesses, which in the worst case are directly counterproductive. It is essential to include an evaluation of the care process regarding these goals.

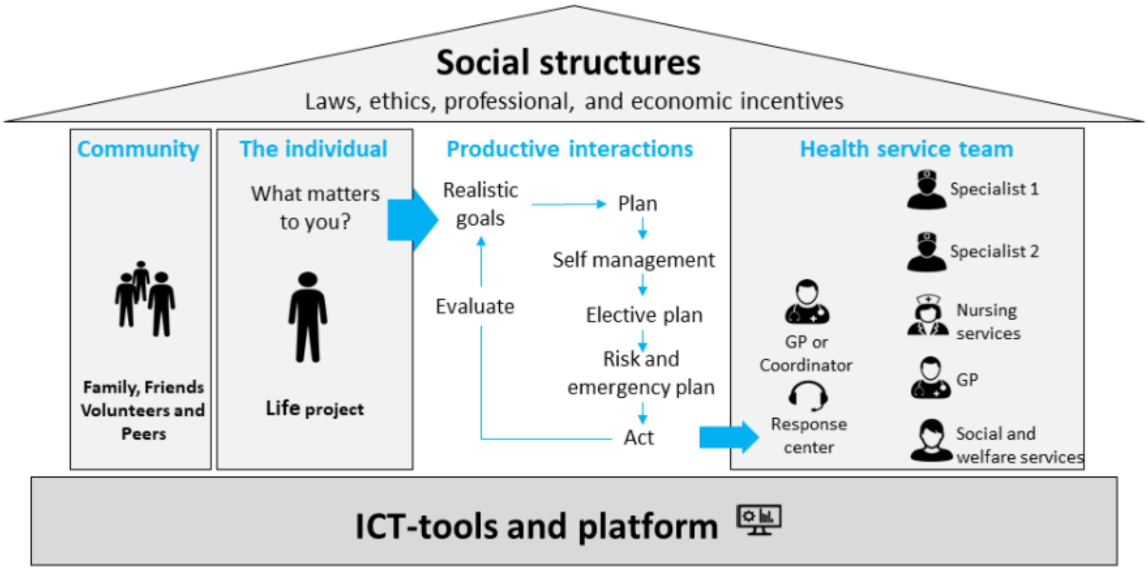
Digital support: (1) tools that help persons define their goals, such as digital access to pertinent information, shared decision-making tools and interactive self-help tools to think about priorities and goals. Included here is low threshold digital communication with providers, (2) applications that record and share *what matters* and the negotiated goals for care, and (3) interactive tools for a shared evaluation of care according to negotiated goals.

Integrated Care

Defining characteristics: We define integrated care as a care plan or a multilateral collaboration, which seeks to meet the goals set above, through the coordination of people, information, and physical resources (ie, aids or medications). Care delivery should proceed according to plan. Severe deviations from the plan should lead to a review and reevaluation of the plan, to adjust and set it on course again.

Care component: The involved parties should map out the roles, responsibilities, and tasks of all parties, including the person and his or her informal carers linked to the common goal. Although it is possible to address different diagnoses and challenges in separate partial plans, these must come together in a bigger plan, to ensure feasibility within the person’s life and alignment with the overarching goals set by the person. As far as possible and desirable, the care plan should reflect best practice. Monitoring of the care plan delivery is essential to catch gaps in care or derailed plans.

Figure 1. The Person-centered integrated care quality framework. The walls, foundation, and roof symbolize the structural resources. The cyclical care process in the house center consists of exploring what matters to the person and translating this into relevant and realistic goals for care, which feed into a multi-professional care plan. The care team delivers care according to the plan, which is continuously adjusted according to a patient and professional joint evaluation of goal attainment. See text for further explanation. (illustration inspired by House of care by Angela Coulter).



Digital support: (1) Tools for interdisciplinary collaboration that identify the care team, roles, and responsibilities, and support communication in virtual teams, including the patient; (2) Tools that support development and sharing of the plan such as decision support and interactive building and update of care plans; (3) Tools that monitor care plan delivery, with alarms in case of critical gaps in care.

Proactive Care

Defining characteristics: Care which seeks to prevent avoidable outcomes that are costly in both human and economic terms, whenever possible and reasonable.

Care component includes both practices that may stabilize and prevent increased risk, such as a comprehensive geriatric assessment [44] and self-management support [45]. It also comprises analyses to identify, monitor, and manage risks factors where early management may prevent clinical deterioration [46]. In population risk management, the idea is to regularly identify persons who might benefit from early intervention [47,48].

Digital support is widely expected to transform proactive care. Improved sensor technology and artificial intelligence are promising better ways to detect risk development, support decision making for both patients and professionals in the face of risk, and finally to provide tools for the evaluation of risk intervention [49].

In summary, Digi-PIP care seeks to reduce the risk of clinical crises that are costly in both human and economic terms, mainly through a digitally supported proactive dimension. However, it cannot be stressed enough that Proactive care will fail, if it is not well enough supported by the people who produce the patient pathway: the patient, his or her significant others, and the care professionals. Person-centered care is necessary to understand and integrate the personal agenda into the care plan. The Integrated care elements are necessary to create the conditions in which the relevant competencies are brought together to design the whole person care plan. Only when these 2 elements are in place, can proactive care across conditions and personal agenda be integrated successfully into the care plan activities. The digital support functions as a change agent and is essential to scalability. The triple aim is a product of the synergies of the 3 PIP elements and their digital support.

Methods

Theoretical Approach

This paper employs a methodological combination of a scoping review and a systematic intervention review of health service and eHealth research literature. The research literature is the most formal arena for new ideas, discussions, and evaluations of current work toward improved care quality for persons with complex long-term needs.

A scoping study approach helps rapidly identify gaps in existing literature relative to a predefined set of expectations, such as a model or acknowledged challenges in the field and points out areas worthy of further attention [50-52]. A systematic intervention review “...attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to assess the benefits and harms of interventions used in healthcare and health policy” [53].

The Search Strategy

The search strategy was set out in a protocol document, outlining the intention of the review, the inclusion and exclusion criteria, and the selection and data extraction methods. We give a more detailed outline of our methods, including exact search terms in Multimedia Appendix 1. We knew from previous searches that we were at risk of few included papers. We, therefore, intentionally made inclusion criteria as broad as possible. We searched Ovid MEDLINE, Web of Science, and Scopus. Publications were eligible if they met the following criteria:

- Target population: must include at least a subset of elderly over 60 years with complex care needs. An author’s description of the study population as being frail, multimorbid, or having complex needs was considered sufficient for fulfilling this criterion. We understand frailty in this context as a state of increased biological vulnerability resulting in a reduced ability to cope with stressors [54].
- Intervention: includes all 3 elements defined as:
    i. Person-centered care: Paper describes person-centered care as an ideal for care, no definition required.
    ii. Integrated care: Either a shared care plan or a multidisciplinary team responsible for the cohesive planning and delivery of care or both.
    iii. Proactive care: early identification of risk, or prevention of risk development at the population or individual level, including self-management support.
- Digital support: any digital technology, supportive of the above intervention components beyond basic electronic health record (EHR) functionality, published after 2000 (to avoid outdated technology).
- Outcome: patient or professional qualitative experience or quantitative measures of the triple aim: (1) patient experience, (2) health outcomes, or (3) cost or benefit ratio.
- Study design: any qualitative or quantitative design, which includes comparisons between situations with or without access to the intervention in either a before-after design or a group comparison.

Exclusion criteria were as follows:

1. Papers that were not original research or had no comparative elements.
2. The study population did not include patients with complex long-term needs.
3. The technology support did not target the Person-centered, Integrated, or Proactive elements. We excluded interventions using a phone, documentation in a general EHR, or shared paper records.

We finalized the search in November 2017. After removal of duplicates, the search identified 927 potentially relevant publications. Moreover, 2 reviewers (KN and KL) independently identified papers that matched the inclusion criteria based on title and abstract. We included 65 publications in the full-text review conducted by FS and GB, and we resolved conflicts between reviewers through discussion until consensus. The study flowchart in Figure 2 shows a final inclusion of 10 publications, originating from 7 studies [55-64].

Extraction of Data and Analyses

The studies included consist of a sociotechnical intervention expected to improve the individual patient pathway and their corresponding evaluation. Each study, which is our unit of observation, may be described by several papers as shown in Table 1. We used Ritchie’s method for applied policy research, which departs from an a priori thematic framework, inspired by either a policy framework or concepts defined in prior work. In this case, the a priori set of concepts were the Digi-PIP care framework described above. The authors code and review the current text, looking for examples, confirmations, or contradictions to these prior themes [65]. We then add other themes that appear to be central to meaning and interpretation to the framework. We produced a condensed description of each intervention, structured by the PIP elements and their digital support, while also describing other themes central to each paper.

In a secondary, deductive-inductive analysis ad modum Tjora [66], we contrasted the ideas of the primary Digi-PIP care with the intervention descriptions. For each of the PIP care elements, we looked at both care innovation and their parallel digital support. We mapped the presence or absence of capabilities to provide Digi-PIP care and compared and contrasted the subcategories to create mutually excluding key components. We revised the mapping to improve clarity of descriptions and to ensure that we had covered both the theoretical and empirical material. The resulting set of key components reflect ideal maturity when all components are present. Finally, we mapped these key components to a matrix according to their contribution in the generic patient pathway stages: Goals, Plans, Delivery, and Evaluation and by their main active ingredient.

Figure 2. Flowchart of a systematic search and inclusions and exclusions of studies of digitally supported person-centered, integrated, and proactive care (Digi-PIP care) for frail multimorbid elderly. Search finalized in November 2017.

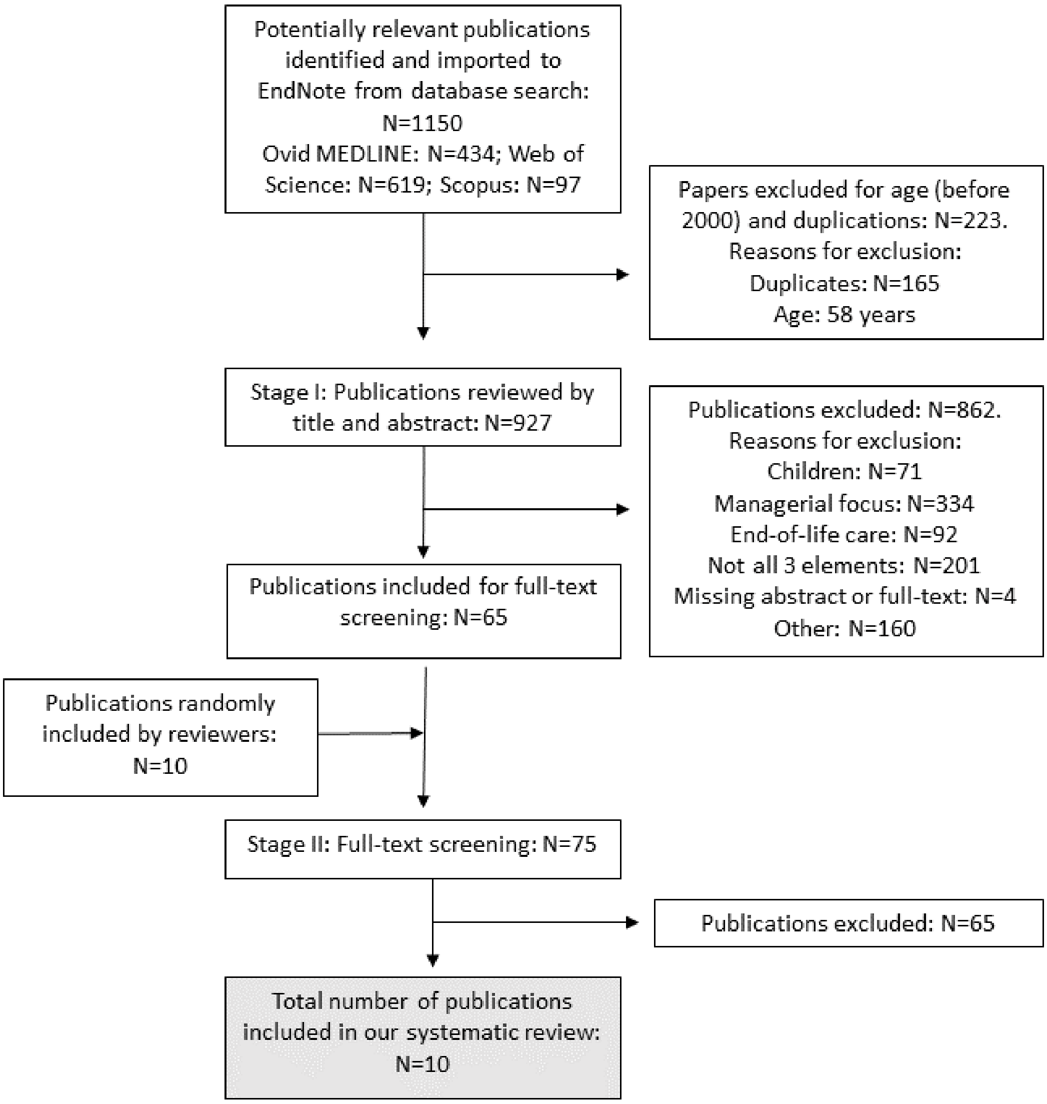


Table 1.

Included papers' authors, publication year	Supporting papers	Acronym or short name	Study context	Study population characteristics
Bleijenberg 2016 [55]	Bleijenberg [67]	U-PROFIT (Utrecht PROactive Frailty Intervention Trial)	Netherland, primary care	Randomized controlled trial (RCT) of frail individuals aged >60 years, using screening tools
Blom 2016 [56]	—[a]	ISCOPE (Integrated Systematic Care for Older People)	Netherland, primary care	RCT of Individuals aged >75 years reporting issues in at least 3 of 4 domains in a screening questionnaire
Martin 2012 [63]	—	PaJR (Patient Journey Record system)	Ireland, primary care	At least 1 chronic condition and high health care use last year randomly allocated in intervention and control groups
Boult 2013 [57], 2011 [58]; Boyd 2008 [60], 2007 [59]	Giddens [68]	Guided care	United States, primary care	Cluster randomized trial of persons > 65 years of age and identified as potential high resource users in screening
Council 2012 [61]	—	PCCP (Person-Centered Care Plan)	United States, Dartmouth, clinic with both primary care and hospital services	5 heuristically selected patients with complex care needs. Before after comparison.
Nelson 2014 [64]	Rosland [25], Kearney [69]	PCMH-VA (Person-centered Medical Home, Veterans Health Administration)	United States, Veterans Health Admin, both primary care and hospital services	Observational study of all patients in the Veterans Health Admin system, with subanalyses for persons with chronic conditions
Liss 2011 [62]	Reid [70]	PCMH-GH (Person-centered Medical Home, Group Health)	United States, group health, primary care	Adults with diabetes, hypertension, and/or coronary heart disease at a Patient Centered Medical Home prototype site compared with other sites in Group Health

[a] The paper had no supporting references relevant to this study.

Assessment of Maturity of the Care System

In an explorative exercise, we used the key components attributed to the 3 active ingredients and their digital support to score each included study regarding maturity. The process of scoring revealed that the interpretation of key components was challenging. The high-level terminology was interpreted differently across authors. Although we were able to arrive at a consensus through dialogue and reflection, it is clear that our scoring system will probably meet the same interpretational challenges in other contexts. Our scores are presented, somewhat tentatively, in Multimedia Appendix 1.

In the next stage, we created the matrix framework, where we translated key system components into system capabilities. A capability is a system's ability to achieve a desired goal or result and does not specify how the task is solved. These capabilities were then mapped to patient-pathways stages so that it would become more evident what we expect from the system at each stage. We hope this matrix may prompt other researchers to engage in reflection and dialogue about the system capabilities necessary to support a PIP care system.

Study Design, Quality, and Effects

We used a best-evidence synthesis approach ad modum de Bruin [16] to summarize both study quality and outcomes. Details of the study-quality evaluation can be found in Multimedia Appendix 1. We excluded studies with 2 or fewer quality points in the quality assessment from the outcome's summary.

Although most of the outcomes were negative, to be parsimonious in our presentation, we have chosen to present all primary outcomes and any secondary outcomes that show a significant effect of the intervention. We present only analyses that are adequately adjusted for baseline biases. We did not attempt to perform a meta-analysis due to the heterogeneity of the contexts, interventions, methodologies, and outcomes. If the intervention reported effects in more than 1 paper, we used data from the latest study.

A short version of the review is presented in Multimedia Appendix 2.

Ethics and Privacy and Registration of Review

This paper is a literature review. It includes no original information on patients, and there are, therefore, no privacy or ethics concerns requiring board review.

Results

Overview of the Empirical Material

Approximately 1.0% (10/927) of the papers identified in the digital search (10 papers representing 7 interventions of 927 potential papers) were included in the final review. We present the intervention context and study population in Table 1.

How Are Digitally Supported Person-Centered, Integrated, and Proactive Care Interventions Operationalized?

Person-Centered Care

All studies claimed to adhere to and acknowledge the long-term aspect of the care process. The degree to which they document that personal goals make an impact on care decisions varied from the mention of Person-centered care as an ideal for care at one end of the scale to a documented impact on care plans and evaluation at the other.

Only the Person-Centered Care Plan (PCCP) study reached the highest possible maturity score on Person-centered maturity. They document how they understand person-centeredness as a journey undertaken by the patient and the surrounding team. The goal in the PCCP is "...to create negotiated goals that incorporate the values of the patient and the healthcare team into a mutually agreed upon explicit action plan." Council demonstrates how the PCCP identifies the care team and distributes responsibilities for goals and tasks, including the goals for which the patient is responsible. The care plan outlines what to do in an emergency. The PCCP is digital, which means it is interactive, updated, and shared among those providers that have access to the same EHR [61]. It is not clear if the patient has access to the PCCP.

Integrated Systematic Care for Older People (ISCOPE) mentions goal setting together with the patient: "...care should be provided proactively to set and prioritize goals together with the patient and to empower the patient to reach these goals." [56]. However, ISCOPE describes the care plan as follows: "Each patient received a geriatric assessment, a comprehensive care plan, evidence-based primary care with proactive follow-up of chronic conditions, coordination of the efforts of health professionals across all health care settings, and facilitated access to community resources." [56]. This description leaves open how the patient goals were incorporated or affected the plan.

The other 5 studies described Person-centered care in terms of patient involvement and engagement. However, they did not document how *what matters* nor how patient needs, values, or preferences were linked to planning and decision making [13,55,57,62-64]. There were no systems that offered digital tools for goal setting, such as shared decision making. No papers mention patient's evaluation of goals or patient-reported outcome measures (PROMs), neither on paper nor digitally.

Integrated Care

Care Plan and Care Manager or Team

All the papers highlight fragmentation of care as a significant challenge and explain how their intervention supports seamless care. Evidence-based and shared care plans are core to the integrated care effort [55-57,61,62]. Digital tools to make the plan, such as evidence-based decision support and health maintenance reminders, were available in the Person-Centered Medical Home, Group Health study [62]. No study mentioned disease-specific paper-based or digital standardized care pathways as a building block for personalized care plans. The PCCP, ISCOPE, and the Person-Centered Medical Home, Veterans Health Administration (PCMH-VA) studies shared the care plan electronically with other providers in a common EHR [57,61,64]. However, it is unclear to which degree the plan is available electronically to parties outside the organization, such as the patient.

In addition to a care plan, 5 of the studies dedicated extra resources to the development, delivery, recruitment of external resources, and follow-up of the care plan [55-60,62,64]. This could take the form of *a case manager* [55-57,63], or a broader, multidisciplinary, *primary care team* [56,57,62]. Except for shared care plans, as described above, we could not find mention of digital support for team communication, such as group video meetings or asynchronous group chats.

Care Delivery

Effect of a care plan is contingent on its actual delivery, but no system monitored care plan delivery systematically. Moreover, 2 studies observed patient progress "by monthly monitoring of symptoms and adherence [57] and measurement-based assessment of progress with facilitation of treatment changes" [69]. Council's PCCP shows both who is responsible for a given action and when it is expected to be done [61]. We found no mention of how care plans were coupled to resource allocation, neither organizationally nor digitally. We found no study where workflow optimization facilitated the transition from planning to delivery of care in any study.

Proactive Care

There are 2 main approaches to proactive care: self-management support and risk and emergency management.

Self-Management Support

Self-management support improves the person's capacity to maintain health and well-being, and it simultaneously strengthens self-agency. Both the PCCP and the Guided Care studies describe how the care plan includes those actions that the patient is responsible for [57,61]. The PCCP study mapped patient strengths and used this as a basis for the self-management support plan. Self-management support also consisted of motivational interviews, patient education, activation workshops, and behavioral counseling [57,62,64,69].

A total of 3 papers did not report any activities on self-management support [55,56,63]. The interventions do not mention widely available self-help applications or self-help communities as tools for self-management. No papers linked patients to community resources.

Risk and Emergency Management

Risk assessment strategies identify and act upon early impactable risk, rather than wait for the clinical crisis. Risk assessment at the individual level was an element in 6 of the 7 studies. The Patient Journey Record system (PaJR) study used lay care guides to keep in touch with patients at high risk for hospitalization. After each conversation, a natural language analysis of the written synopsis would estimate the risk of hospitalization [63]. Digital health analytics from GP-EHR or insurance data identified frailty or risk for high care expenses in the Utrecht PROactive Frailty Intervention Trial (U-PROFIT) [55] and Guided care studies [57], and, in the latter study, also regular follow-up phone calls to assess changes in risk [57]. In the PMCH-GH intervention, patients filled an eHealth risk assessment form [62]. In the ISCOPE study, the risk was assessed once through a questionnaire sent to elderly citizens, without the use of digital tools. In the PMCH-VA study, they mapped unhealthy lifestyle habits [64]. Only the PCCP study had no risk assessment included [61].

The risk identification was linked to a range of *actions*. In the PaJR study, the lay care guide would notify the appropriate usual care service, such as the general practitioner or a case manager [63]. In the U-PROFIT [55], ISCOPE [56], and Guided Care studies, a nurse or GP-nurse team would develop a comprehensive and evidence-based geriatric care plan [55-57]. The PMCH-VA self-reported risk assessment is not described to be directly linked to any intervention activity, except for general training in lifestyle coaching and motivational interviewing to all clinical employees [25,69]. Remote telehealth follow-up from registered nurses was mentioned as an opportunity in some VA facilities [25]. No studies described digital decision support for alarm situations.

Only the PCCP plan summarized the emergency measures agreed by both patient and providers [61]. Low-threshold one point of contact, such as virtual video contacts for emergencies and questions, was emphasized in the VA and Group health [62,64]. Other papers outlined no emergency contingency plans [55-57,62-64].

The Maturity of Digitally Supported Person-Centered, Integrated, and Proactive Care Models

We have conceptualized Digi-PIP care as a set of characteristics systematically included in the individual Patient Pathway in support of the 4 generic stages of a patient pathway [41]. Each key component is described in terms of the capability the system offers. The more elements present, the higher the maturity of the care system in question (see Table 2).

Although all the included papers addressed the 3 active ingredients and some form of digital support, it was clear that none of the interventions succeeded in giving equal focus to all elements.

The digital support, particularly, was marginal and far less advanced than what is considered state of the art in research projects addressing only 1 of the PIP digital axis. PCMH-VA had the highest total digital score, with a patient portal, low-threshold e-visits, a shared care plan, and telemonitoring services. They were far from sporting a full suite of eHealth services that would both leverage and scale their PCMH approach. All other studies lacked digital support in at least 1 PIP elements. We believe that a genuinely sociotechnical design of PIP care, where technology supports and replaces analog services, is still somewhere in the future.

Even when the care packages are complex, the studies do not explicitly acknowledge the interwoven dependencies between the PIP components. There is some understanding that all Digi-PIP ingredients are essential. For instance, in U-PROFIT study, the main focus was on integrated care, whereas the person-centered and proactive components were considered integral parts of the overarching comprehensive care plan [55].

However, the focus on the unbroken chain of care events that lead from intervention to the desired outcome is not always present. For example, care outcomes require loyalty in care delivery to the care plan. Impressive as the service redesign efforts in the Guided Care study is, the Guided care nurse is *grafted* onto the usual care system in a supportive role. If the carefully crafted Guided care plan does not fit with the agenda of the usual providers, the guided care plan may be set aside [59]. In the VA study by Nelson, care professionals were encouraged to explore *what matters* to patients and used motivational interviewing to do so. The integrated care planning did not seem to build on the goals brought forward by the patients. Moreover, the proactive component included mostly primary preventions such as smoking cessation but did not include individualized risk monitoring and management, which is presumably more efficient. The clever monitoring of a high-risk situation is only effective if the response to deal with that risk is adequate [25,69]. We found similar weak links in the *chain of success* in all of the 7 studies.

Table 2. Key care and digital components, described in terms of the capabilities they offer in support of the person-centered, integrated, and proactive care. We have mapped each PIP-element to statements of care system capability for each of the 4 generic stages of an individualized patient pathway.

Care components	Goals	Plans	Delivery	Evaluation
The person-centered care system…				
Care	...declares Person-centered care as an ideal and explores “what matters to me?” and “patient values, needs, and preferences.”	...uses “What matters to me?” to negotiate realistic goals and create a care plan.	...includes patient capabilities aligned with “What matters to me?” in care delivery.	...asks for patient feedback/ PROMs[a]
Digital support	...offers access to digital health information/ electronic health record and supports the formulation of “what matters to me?”	...offers digital sharing of: What my carers should know about me.	...includes the patient in virtual care delivery and team exchanges.	...encourages digital feedback from patients, including PROMs.
The integrated care system…				
Care delivery	...identify condition- or function-specific goals that support “what matters.”	...combines condition/function-specific pathways into a whole person care plan for all conditions.	...allocates resources to care plan, to show who does what when.	...follow up to identify needs for adjustment of care plans or delivery.
Digital support	...digitally identifies potential professionals to contribute to care plan development aligned with “what matters to me?”	...provide tools to build a personalized digital evidence-based care plan, with workflow optimization to show: who does what when.	...shares the care plan digitally across providers and offers tools for virtual team communication (video, messages, and chat).	...triggers an alarm in case of gaps in critical care delivery.
The proactive care system…				
Care delivery	...identifies high-risk subpopulations, their individual high-risk scenarios over time and aligns focus on risk with ”What matters to me?“	...supports risk monitoring, self-managed or professional follow-up, in alignment with ”What matters to me?“	...offers low threshold response (self-managed, office or home visits) to uncertainties, emergencies, and alarms.	...learns and adjusts goals and plans in light of undesired events and ”What matters to me?“
Digital support	...offers an algorithm-based risk-stratification tool to identify high-risk populations and their individual risk scenarios over time.	...offers personal digital health apps and sensors that monitor risk and provide digital contingency plans in case of uncertainty, emergencies, or alarms.	...provides digital decision support and low-threshold e-visits in case of uncertainty, emergencies, or alarms.	...is a learning health care system improves prediction and action plans in light of undesired events.

[a] PROMs: patient-reported outcome measures.

Study Design Quality and Effects

In compliance with de Bruin’s methodology [16], we include only studies with a quality score of 3 or more in our effect report, which left us with 4 studies. We present the quality scores in Multimedia Appendix 1. The included studies are 3 cluster randomized trials at General practice level (U-PROFIT, ISCOPE, and Guided Care) reporting on patient-level data, whereas the fourth is an observational study utilizing aggregated measures at clinic level in the PMCH-VA study.

The U-PROFIT, ISCOPE, Guided Care, and PCMH-VA studies present 8, 3, 7, and 6 outcome analyses, respectively, 24 in all. Outcomes lie within areas of patient satisfaction, quality of life, function, disease process quality, health care utilization, mortality, and staff burnout. Of these, only the emergency department visits in Nelson’s study from the VA and the home-care visit frequency in Guided Care showed clear and clinically meaningful significant positive effects (see Table 3) [57,64]. The U-PROFIT study showed a slower functional decline in the intervention group compared with controls, but the clinical significance was deemed uncertain (Table 3) [55]. None of the other analyses showed significant effects of the intervention.

This study cannot answer the question of whether intervention maturity matters, because the number of high-quality studies is not large enough to support a correlation analysis between maturity and outcomes.

Table 3. Selected outcomes in 4 high-quality studies of *digitally supported Person-centered, Integrated, and Proactive care* (digi-PIP-care) for frail multimorbid elderly. All primary outcomes and any positive secondary outcomes analyses are shown. Negative secondary analyses not presented.

Paper	Outcome measure	Patient or clinics	N	Effect intervention	Effect control	Ratio Intervention/ Control	*P* value
Utrecht PROactive Frailty Intervention Trial [55]	Katz 15 scores at 6 months. Range 0-15, lower score is better	P[a]	2754	1.7	1.7	0.97	Not significant
	Katz 15 scores at 12 months. Range 0-15, lower score is better	P	2489	1.9	2.0	0.92	.03
Integrated Systematic Care for Older People [56]	12 months follow-up, change in quality of life, Cantril's ladder (range 0-10, higher is better)	P	842	−0.2	−0.2	1.00	.82
	12 months follow-up, Delta Groningen Activities Restriction Scale (range 18-72, lower score is better).	P	842	2.9	3.5	0.83	.30
Guided Care [57]	Functional health Short Form 36, higher score is better	P	477	36.1	37.5	0.96	Not significant
	Home health care episodes	P	477	0.9	1.3	0.71	<.05
Person-centered Medical Home-Veterans Health Administration [64]	Emergency Department visits per 1000 patients per year (secondary outcome)	Clinic	913	188	245	0.77	<.001

[a] P: Patient.

Discussion

Summary of Findings

We identified 927 potentially eligible papers, but after full-text review, we included only 10 papers.

The PIP elements were supported to a varying degree. Person-centeredness was an ideal for care, but only one intervention made *what mattered to the person* visible throughout the care plan. The studies counter care fragmentation through a whole-person care plan and the engagement of care-coordinating case managers or multiprofessional teams. Care delivery according to plan seems to be taken for granted as none monitor system delivery of care plans. Although papers mention self-management support and emergency plans are mentioned, risk identification and management are the main proactive strategy. No studies suggested workflow optimization or patient-reported outcomes.

The most prominent digital support of the PIP elements was risk stratification tools. Second, 3 providers supported the sharing of care plans in the EHR.

The chain of care is only as strong as its weakest link. The maturity matrix made it possible to identify several potential breaks or weaknesses in the chain of success. The most common weaknesses were:

1. A lack of documentation that *what mattered to the person* was also brought to bear on care plans and delivery.
2. That planned care plan was actually delivered.
3. That risk identification schemes were coupled with adequate responses from the care system.
4. Finally, there are feedback loops that support learning and adjustment of the PIP dimensions of care.

We included 4 studies in our summary of effects after methodological quality assessment. Moreover, 2 of the 24 analyses in 4 studies reported modest positive outcomes with reductions in emergency care utilization and home health visits.

The Invisible Sociotechnical Care Process

The specialization of health care remains both its biggest asset and weakness. Systems theory has shown long ago that when systems grow, they tend to specialize. If specialization is not coupled with centralized coordination, the organization's ability to deliver its end product will be crippled [71]. In other services (banking, the tourist industry, publishing, and e-commerce), digital tools are the glue that allows all the involved parties, irrespective of professional and organizational affiliation, to work effectively together. The digitally supported processes organize people, information, and things into a value chain for the patient. The need to share goals and plans, understand roles and tasks, and learn to support each other collectively is the same in health care as in other industries. The digital infrastructure is lagging. Strict privacy rules, lack of e-governance structures, lack of interoperability standards, and lack of business models for eHealth vendors are acknowledged barriers [72]. The infrastructure is slowly coming into place, but not nearly fast enough to catch up with other industries.

Weaknesses in the Chain of Success

We structured this review according to the Digi-PIP care framework and a maturity evaluation. As noted, there is an abundance of literature and reviews supporting each of the PIP elements alone [15-17,31,45,49]. However, all of the reviews also note the heterogeneity of the interventions and the lack of consistency in results. In this light, we hypothesize that the wide variation of effects in Digi-PIP care interventions are attributable to the weaknesses in the care system's implementation and understanding of *chains of success*: The current efforts are not bringing us closer to the triple aim. So what needs to change? We claim that approaches to date have failed to address the full complexity of the problem, both in the health care system design itself and in the corresponding scientific intervention with outcome analysis approaches.

Complex Adaptive Systems

Scientists are taught to narrow down and examine *one factor at a time*, and the randomized controlled trial (RCT) is the golden standard of how to ascertain the effect of the single factor. RCTs are designed for hypothesis testing, that is, A works on B to produce C with mechanism D, when context X, Y, and Z are stable. However, this requires the researcher to have a reasonable hypothesis on how the *one factor* works, under what conditions it works, and to keep all other conditions stable for the duration of the experiment. The RCT has, however, become a *test tool* for complex interventions in complex settings, where the assumptions under scrutiny are unclear. The validity of such an RCT is low because there is no clear hypothesis. If we test an antibiotic on cases of viral infection, it will look as if the antibiotic did not work. The truth is that we did not understand the mechanism of action well enough to design the study correctly. We need to move beyond the RCT and use other tools to understand how to improve the outcomes of the complex and fluid social processes of health care.

A Good Enough Vision

Complexity theory [73] and quality improvement theory [74,75] prescribe a different set of methods to understand, improve, and predict outcomes in complex adaptive systems [76]. Although it is outside our scope to describe complex adaptive systems in full, there are some important points worth reflecting over in light of our review. Complexity theory predicts that a linear plan where method X leads to outcome A will not be successful in a complex adaptive systems. This is because unknowable factors will frustrate the method X in a proportion of the cases.

Instead of placing all faith in method X, the intervention will need to include a way to detect and manage challenges as professionals become aware of them. In complex adaptive systems, one moves toward a goal first and foremost by creating a good enough vision of what the goal is. In Digi-PIP care, we believe that vision is a negotiated and realistic set of personalized goals aligned with *what matters* to the person. The next step is to provide frontline professionals with an array of tools and checkpoints that are useful in the creation of an individualized road map toward the personalized goals. In Digi-PIP care, this will be a set of professional knowledge and skill sets, more or less evidence-based, more or less proactive, assembled in the integrated care team. Depending on the context, the agents must apply their knowledge and experience to choose a way forward. If they are closing in on the goal, they stay on course. If the goal is slipping, they must reevaluate and adjust. In Digi-PIP care, that will amount to evaluation not only of the person’s overarching goals, which may be long term and difficult to maneuver by. Choosing proximal and sensitive subgoals that support the long-term goals may also be useful. The capabilities identified in the maturity matrix may serve as tools to create subgoals in each case. Adopting a *chain of success* way of thinking to understand which factors must line up to create a desired outcome in the unique case can be productive. For instance, the lack of a care plan is an obvious impediment to seamless, coherent care delivery. However, a care plan does not translate into care delivery by magic. Assuring care delivery according to plan is a part of the challenge to reach the desired effect. A close dialogue with the patient throughout the process will help this continuous guiding evaluation to take place. There also needs to be a sensitivity to changing goals, as insights and contexts change for the patient and the team.

Improving Outcomes in Complex Adaptive Systems

The researchers who work in improving outcomes in complex adaptive systems will be working with questions such as How can we help formulate good enough goals and subgoals, that are also observable, and provide process guidance?; What are areas of knowledge, skills, and tools essential to make available to enable professionals to invent proper processes?; Are there standardized components that can be plugged into and tailored to the individual pathway?; and How can we help frontline professionals hypothesize and modulate the chain of events that will lead to success in each case?

The Validity of Our Work—Strengths and Limitations

We have done systematic searches in the 3 largest literature databases covering the health and eHealth field. A librarian trained in building complex searches conducted the search. Pairs or triplets of authors performed all the steps in the screening of papers, data extraction, and grading of papers. Author-pairs discussed disagreements until they reached consensus, or, if it concerned a matter of principle, the entire author group addressed the issue. We defined rather broad inclusion criteria so that we should not inadvertently exclude studies that might bring forth new knowledge. The authors were, with one exception, senior researchers at the professor or associate professor level with long track records in the health and eHealth fields. These are the strengths of the study, which ensure that we have indeed identified the current published knowledge base regarding Digi-PIP care.

Given the extensive activity in this field and the many large enterprises underway in this area, we were surprised by the meager catch and by how old many of the papers were. We believe that the small number of articles indicates that this is an area that many researchers find too complicated to bring into a viable research model. Those organizations that are making progress in this field do so without publishing their interventions, the digital solutions, and their results. There may well be significant experiential knowledge in the field that we do not capture in this review.

Conclusions

The research literature is permeated with *common sense* argumentation that each of the Digi-PIP elements is important. However, many hardworking and extremely talented health care scientists have to date chosen to address just a subset of the elements in their research. We identified only 10 papers from 7 interventions reporting on studies encompassing all elements together, and they all report a limited effect. The general lack of effects on the triple aim from both subset approaches and all element approaches until now is also disappointing for the research and practice field. We argue that it is now time to rethink our approaches to health care innovation by acknowledging the patient voice and the inherent system complexities.

It is not enough to provide a care plan that seems sensible to the providers. It must also be *owned* by the person himself or herself, who is the crucial resource, enabler, and guide. It is not enough to make a care plan unless there is also considerable devotion to the delivery of the care plan and attention to whether its objectives are met. It is not enough to be proactive in an elective care plan if one does not also monitor risk for the impending crisis and provide an emergency plan, which can help avoid it. Finally, it is probably not possible to address these complex processual challenges with regular EHR, paper, and phone. It is time to say that the health care sector is under-digitized and that the lack of appropriate digital support is a barrier for PIP-care and costs both patients and professionals dearly.

We believe that a reductionist scientific methodology is blocking the way forward. We need to embrace the problem-solving methods suited for the improvement of outcomes in complex adaptive systems. Researchers need to embrace questions such as How do we formulate good enough process guiding goals?; How does a professional formulate a rationale for adjustment of a process?; and What are the generalizable components of *individualized* pathway creation?

We predict that research will not show consistent results from care transformation for persons with complex long-term needs until all 3 PIP care elements are successfully and flexibly implemented with digital support. We need a chain of success thinking in the work of creating patient pathways. The art of high-quality care is to invent a road as it is being walked, toward *what matters to the person* in every pathway for patients with complex long-term needs.

Acknowledgments

The authors are grateful for the competent and timely help of the research librarian Karianne Lind (KL) in performing database searches and sifting through titles and abstracts in search for eligible papers.

The study was funded by 2 sources: (1) The Norwegian Research Council through the m&e Health project (contract 269884/H10) and (2) The Norwegian Centre for E-health Research. This work has been undertaken in complete independence from funders. The publication charges for this study have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Authors' Contributions

The authors represented a set of multiprofessional competencies: technological/eHealth (FS, MR, RF, and BS), medical (GB), and nursing (BS, KN) disciplines, and they are all experienced health service researchers. Some of the authors also had patient experience. GB and FS were responsible for writing the protocol. KN and our research librarian (KL—see Acknowledgments) performed searches and initial screening. The first author was responsible for drafting the main text. All authors took part in the reading of original literature, extraction of data and data analysis, critical review of drafts, and approved the last version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

More detail on methods, and literature review search strings.

[DOCX File, 132KB - jmir_v21i4e12517_app1.docx]

Multimedia Appendix 2

The short version: The evidence base for an ideal care pathway for frail multi-morbid elderly: A combined scoping and systematic intervention review.

[PPTX File, 1MB - jmir_v21i4e12517_app2.pptx]

References

1. Afshar S, Roderick P, Kowal P, Dimitrov B, Hill A. Global Patterns of Multimorbidity: A Comparison of 28 Countries Using the World Health Surveys. In: Hoque MN, Pecotte B, McGehee MA, editors. Applied Demography and Public Health in the 21st Century. Switzerland: Champringer International Publishing; 2017:381-402.

2. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, et al. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. PLoS One 2014;9(7):e102149 [FREE Full text] [doi: 10.1371/journal.pone.0102149] [Medline: 25048354]
3. van Oostrom SH, Gijsen R, Stirbu I, Korevaar JC, Schellevis FG, Picavet HS, et al. Time trends in prevalence of chronic diseases and multimorbidity not only due to aging: data from general practices and health surveys. PLoS One 2016;11(8):e0160264 [FREE Full text] [doi: 10.1371/journal.pone.0160264] [Medline: 27482903]
4. Wang L, Si L, Cocker F, Palmer AJ, Sanderson K. A systematic review of cost-of-illness studies of multimorbidity. Appl Health Econ Health Policy 2018 Feb;16(1):15-29. [doi: 10.1007/s40258-017-0346-6] [Medline: 28856585]
5. Zulman DM, Pal CC, Wagner TH, Yoon J, Cohen DM, Holmes TH, et al. Multimorbidity and healthcare utilisation among high-cost patients in the US Veterans Affairs Health Care System. BMJ Open 2015 Apr 16;5(4):e007771 [FREE Full text] [doi: 10.1136/bmjopen-2015-007771] [Medline: 25882486]
6. Heiberg I. High utilisation patients in somatic specialist healthcare in Northern Norway [Storforbrukere av somatisk spesialisthelsetjeneste i Helse Nord. SKDE]. Tromsø, Norway: www.skde.no; 2015:1-63.
7. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
8. Conway M. How do committees invent? Datamation 1968;Vol 14(4):28-31 [FREE Full text]
9. Coiera E. Four rules for the reinvention of health care. Br Med J 2004 May 15;328(7449):1197-1199 [FREE Full text] [doi: 10.1136/bmj.328.7449.1197] [Medline: 15142933]
10. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769 [FREE Full text] [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
11. Patient-Centered Primary Care Collaborative.. Patient-Centered Primary Care Collaborative. 2015. Patient-Centered Medical Home 201: A Snapshot of the Evidence URL: https://www.pcpcc.org/resource/map-tools-pcmh-slide-presentations [accessed 2018-10-16] [WebCite Cache ID 73DIUAtHF]
12. World Health Organization. 2016. Framework on integrated, people-centred health services. Report by the Secretariat URL: http://apps.who.int/gb/ebwha/pdf_files/WHA69/A69_39-en.pdf?ua=1 [WebCite Cache ID 73DWSHo7t]
13. Rich E ELM, Lipson D, Libersky J, Parchman M. Patient-Centered Medical Home. 2012. Coordinating Care for Adults With Complex Care Needs in the Patient-Centered Medical Home: Challenges and Solutions URL: https://pcmh.ahrq.gov/sites/default/files/attachments/Coordinating%20Care%20for%20Adults%20with%20Complex%20Care%20Needs.pdf [WebCite Cache ID 73DWky5eg]
14. Joint Action CHRODIS. 2017. Report on care pathways approaches for multimorbid chronic patients URL: http://chrodis.eu/wp-content/uploads/2018/01/multimorbidity_care_model.pdf [accessed 2018-10-16] [WebCite Cache ID 73DWtuDxj]
15. Coulter A, Entwistle VA, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database Syst Rev 2015 Mar 03(3):CD010523 [FREE Full text] [doi: 10.1002/14651858.CD010523.pub2] [Medline: 25733495]
16. de Bruin SR, Versnel N, Lemmens LC, Molema CC, Schellevis FG, Nijpels G, et al. Comprehensive care programs for patients with multiple chronic conditions: a systematic literature review. Health Policy 2012 Oct;107(2-3):108-145. [doi: 10.1016/j.healthpol.2012.06.006] [Medline: 22884086]
17. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. Br Med J 2012;344:e3874 [FREE Full text] [Medline: 22723612]
18. Hallberg IR, Kristensson J. Preventive home care of frail older people: a review of recent case management studies. J Clin Nurs 2004 Sep;13(6B):112-120. [Medline: 15724826]
19. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
20. Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes from health information exchange: systematic review and future research needs. JMIR Med Inform 2015 Dec 15;3(4):e39 [FREE Full text] [doi: 10.2196/medinform.5215] [Medline: 26678413]
21. Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications. J Am Med Inform Assoc 2016 May;23(3):491-495. [doi: 10.1093/jamia/ocv199] [Medline: 26911810]
22. Ham C, York N, Sutch S, Shaw R. Hospital bed utilisation in the NHS, Kaiser Permanente, and the US Medicare programme: analysis of routine data. Br Med J 2003 Nov 29;327(7426):1257 [FREE Full text] [doi: 10.1136/bmj.327.7426.1257] [Medline: 14644968]
23. Feachem RG, Sekhri NK, White KL. Getting more for their dollar: a comparison of the NHS with California's Kaiser Permanente. Br Med J 2002 Jan 19;324(7330):135-141 [FREE Full text] [Medline: 11799029]
24. Driscoll DL, Hiratsuka V, Johnston JM, Norman S, Reilly KM, Shaw J, et al. Process and outcomes of patient-centered medical care with Alaska Native people at Southcentral Foundation. Ann Fam Med 2013;11(Suppl 1):S41-S49 [FREE Full text] [doi: 10.1370/afm.1474] [Medline: 23690385]

25. Rosland AM, Nelson K, Sun H, Dolan ED, Maynard C, Bryson C, et al. The patient-centered medical home in the Veterans Health Administration. Am J Manag Care 2013 Jul 01;19(7):e263-e272 [FREE Full text] [Medline: 23919446]
26. Busse R, Stahl J. Integrated care experiences and outcomes in Germany, the Netherlands, and England. Health Aff (Millwood) 2014 Sep;33(9):1549-1558. [doi: 10.1377/hlthaff.2014.0419] [Medline: 25201659]
27. Hughes G. New models of care: the policy discourse of integrated care. PPP 2017 Oct 26;11(2):72-89. [doi: 10.3351/ppp.2017.6792867782]
28. Salisbury C, Man M, Bower P, Guthrie B, Chaplin K, Gaunt DM, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. Lancet 2018 Dec 07;392(10141):41-50 [FREE Full text] [doi: 10.1016/S0140-6736(18)31308-4] [Medline: 29961638]
29. Scott IA. Chronic disease management: a primer for physicians. Intern Med J 2008 Jun;38(6):427-437. [doi: 10.1111/j.1445-5994.2007.01524.x] [Medline: 18298554]
30. Beswick AD, Rees K, Dieppe P, Ayis S, Gooberman-Hill R, Horwood J, et al. Complex interventions to improve physical function and maintain independent living in elderly people: a systematic review and meta-analysis. Lancet 2008 Mar 01;371(9614):725-735 [FREE Full text] [doi: 10.1016/S0140-6736(08)60342-6] [Medline: 18313501]
31. Lemmens LC, Molema CC, Versnel N, Baan CA, de Bruin SR. Integrated care programs for patients with psychological comorbidity: a systematic review and meta-analysis. J Psychosom Res 2015 Dec;79(6):580-594. [doi: 10.1016/j.jpsychores.2015.07.013] [Medline: 26354890]
32. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. Int J Integr Care 2013;13:e010. [Medline: 23687482]
33. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. Br Med J 2007 Jul 7;335(7609):24-27 [FREE Full text] [doi: 10.1136/bmj.39246.581169.80] [Medline: 17615222]
34. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26 [FREE Full text] [doi: 10.5888/pcd10.120180] [Medline: 23428085]
35. Huntley AL, Thomas R, Mann M, Huws D, Elwyn G, Paranjothy S, et al. Is case management effective in reducing the risk of unplanned hospital admissions for older people? A systematic review and meta-analysis. Fam Pract 2013 Jun;30(3):266-275. [doi: 10.1093/fampra/cms081] [Medline: 23315222]
36. Tenforde AS, Hefner JE, Kodish-Wachs JE, Iaccarino MA, Paganoni S. Telehealth in physical medicine and rehabilitation: a narrative review. PM R 2017 May;9(5S):S51-S58. [doi: 10.1016/j.pmrj.2017.02.013] [Medline: 28527504]
37. Kaplan B, Harris-Salamone KD. Health IT success and failure: recommendations from literature and an AMIA workshop. J Am Med Inform Assoc 2009;16(3):291-299 [FREE Full text] [doi: 10.1197/jamia.M2997] [Medline: 19261935]
38. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
39. The Health Foundation. 2014. Person-centered care made simple URL: https://www.health.org.uk/sites/health/files/PersonCentredCareMadeSimple.pdf [WebCite Cache ID 73DX9aYtT]
40. Ehrlich C, Kendall E, Muenchberger H, Armstrong K. Coordinated care: what does that really mean? Health Soc Care Community 2009 Nov;17(6):619-627. [doi: 10.1111/j.1365-2524.2009.00863.x] [Medline: 19469914]
41. Berntsen G, Høyem A, Lettrem I, Ruland C, Rumpsfeld M, Gammon D. A person-centered integrated care quality framework, based on a qualitative study of patients' evaluation of care in light of chronic care ideals. BMC Health Serv Res 2018 Dec 20;18(1):479 [FREE Full text] [doi: 10.1186/s12913-018-3246-z] [Medline: 29925357]
42. Coulter A, Roberts S, Dixon A. The Kings Fund. London; 2013. Delivering better services for people with long-term conditions: Building the house of care URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/delivering-better-services-for-people-with-long-term-conditions.pdf [WebCite Cache ID 73DY4ZvtB]
43. Taylor C. The Malaise Of Modernity. Toronto: House Of Anansi Press; 1998.
44. Turner G, Clegg A, British Geriatrics Society, Age UK, Royal College of General Practioners. Best practice guidelines for the management of frailty: a British Geriatrics Society, Age UK and Royal College of General Practitioners report. Age Ageing 2014 Nov;43(6):744-747 [FREE Full text] [doi: 10.1093/ageing/afu138] [Medline: 25336440]
45. McBain H, Shipley M, Newman S. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. BMC Health Serv Res 2015 Dec 18;15:565 [FREE Full text] [doi: 10.1186/s12913-015-1221-5] [Medline: 26684011]
46. Ferdous R, Khan F, Sadiq R, Amyotte P, Veitch B. Analyzing system safety and risks under uncertainty using a bow-tie diagram: an innovative approach. Process Saf Environ Prot 2013 Jan;91(1-2):1-18. [doi: 10.1016/j.psep.2011.08.010]
47. Bakker FC, Robben SH, Olde Rikkert MG. Effects of hospital-wide interventions to improve care for frail older inpatients: a systematic review. BMJ Qual Saf 2011 Aug;20(8):680-691. [doi: 10.1136/bmjqs.2010.047183] [Medline: 21355019]
48. Eklund K, Wilhelmson K. Outcomes of coordinated and integrated interventions targeting frail elderly people: a systematic review of randomised controlled trials. Health Soc Care Community 2009 Sep;17(5):447-458. [doi: 10.1111/j.1365-2524.2009.00844.x] [Medline: 19245421]
49. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]

50. Gammon D, Berntsen GK, Koricho AT, Sygna K, Ruland C. The chronic care model and technological research and innovation: a scoping review at the crossroads. J Med Internet Res 2015;17(2):e25 [FREE Full text] [doi: 10.2196/jmir.3547] [Medline: 25677200]
51. 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]
52. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
53. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 6.: The Cochrane Collaboration; 2011. URL: https://training.cochrane.org/handbook [accessed 2019-03-14] [WebCite Cache ID 76sMUFu3k]
54. Xue QL. The frailty syndrome: definition and natural history. Clin Geriatr Med 2011 Feb;27(1):1-15 [FREE Full text] [doi: 10.1016/j.cger.2010.08.009] [Medline: 21093718]
55. Bleijenberg N, Drubbel I, Schuurmans MJ, Dam HT, Zuithoff NP, Numans ME, et al. Effectiveness of a proactive primary care program on preserving daily functioning of older people: a cluster randomized controlled trial. J Am Geriatr Soc 2016 Dec;64(9):1779-1788. [doi: 10.1111/jgs.14325] [Medline: 27459236]
56. Blom J, den Elzen W, van Houwelingen AH, Heijmans M, Stijnen T, Van den Hout W, et al. Effectiveness and cost-effectiveness of a proactive, goal-oriented, integrated care model in general practice for older people. A cluster randomised controlled trial: Integrated Systematic Care for older People--the ISCOPE study. Age Ageing 2016 Jan;45(1):30-41 [FREE Full text] [doi: 10.1093/ageing/afv174] [Medline: 26764392]
57. Boult C, Leff B, Boyd CM, Wolff JL, Marsteller JA, Frick KD, et al. A matched-pair cluster-randomized trial of guided care for high-risk older patients. J Gen Intern Med 2013 May;28(5):612-621 [FREE Full text] [doi: 10.1007/s11606-012-2287-y] [Medline: 23307395]
58. Boult C, Reider L, Leff B, Frick KD, Boyd CM, Wolff JL, et al. The effect of guided care teams on the use of health services: results from a cluster-randomized controlled trial. Arch Intern Med 2011 Mar 14;171(5):460-466 [FREE Full text] [doi: 10.1001/archinternmed.2010.540] [Medline: 21403043]
59. Boyd CM, Boult C, Shadmi E, Leff B, Brager R, Dunbar L, et al. Guided care for multimorbid older adults. Gerontologist 2007 Oct;47(5):697-704. [Medline: 17989412]
60. Boyd CM, Shadmi E, Conwell LJ, Griswold M, Leff B, Brager R, et al. A pilot test of the effect of guided care on the quality of primary care experiences for multimorbid older adults. J Gen Intern Med 2008 May;23(5):536-542 [FREE Full text] [doi: 10.1007/s11606-008-0529-9] [Medline: 18266045]
61. Council LS, Geffken D, Valeras AB, Orzano AJ, Rechisky A, Anderson S. A medical home: changing the way patients and teams relate through patient-centered care plans. Fam Syst Health 2012 Sep;30(3):190-198. [doi: 10.1037/a0029832] [Medline: 22985385]
62. Liss DT, Fishman PA, Rutter CM, Grembowski D, Ross TR, Johnson EA, et al. Outcomes among chronically ill adults in a medical home prototype. Am J Manag Care 2013 Oct 01;19(10):e348-e358 [FREE Full text] [Medline: 24304182]
63. Martin CM, Vogel C, Grady D, Zarabzadeh A, Hederman L, Kellett J, et al. Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR). J Eval Clin Pract 2012 Dec;18(6):1226-1234. [doi: 10.1111/j.1365-2753.2012.01880.x] [Medline: 22816797]
64. Nelson KM, Helfrich C, Sun H, Hebert PL, Liu C, Dolan E, et al. Implementation of the patient-centered medical home in the Veterans Health Administration: associations with patient satisfaction, quality of care, staff burnout, and hospital and emergency department use. JAMA Intern Med 2014 Aug;174(8):1350-1358. [doi: 10.1001/jamainternmed.2014.2488] [Medline: 25055197]
65. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: The Qualitative Researcher's Companion. Los Angeles, CA: Sage Publications; 2002:305-329.
66. Tjora A. Qualitative research methods - a practical guide [Kvalitative forskningsmetoder i praksis.]. Oslo: Gyldendal norsk forlag AS; 2012.
67. Bleijenberg N, Ten Dam VH, Drubbel I, Numans ME, de Wit NJ, Schuurmans MJ. Treatment fidelity of an evidence-based nurse-led intervention in a proactive primary care program for older people. Worldviews Evid Based Nurs 2016 Feb;13(1):75-84. [doi: 10.1111/wvn.12151] [Medline: 26873373]
68. Foret GJ, Tanner E, Frey K, Reider L, Boult C. Expanding the gerontological nursing role in Guided Care. Geriatr Nurs 2009;30(5):358-364. [doi: 10.1016/j.gerinurse.2009.08.006] [Medline: 19839117]
69. Kearney LK, Post EP, Zeiss A, Goldstein MG, Dundon M. The role of mental and behavioral health in the application of the patient-centered medical home in the Department of Veterans Affairs. Transl Behav Med 2011 Dec;1(4):624-628 [FREE Full text] [doi: 10.1007/s13142-011-0093-4] [Medline: 24073086]
70. Reid RJ, Fishman PA, Yu O, Ross TR, Tufano JT, Soman MP, et al. Patient-centered medical home demonstration: a prospective, quasi-experimental, before and after evaluation. Am J Manag Care 2009 Sep 01;15(9):e71-e87 [FREE Full text] [Medline: 19728768]
71. Bertalanffy L. An outline of general system theory. Br J Philos Sci 1950;I(2):134-165. [doi: 10.1093/bjps/I.2.134]

72. Weinelt B. World Economic Forum White Paper. Digital Transformation of Industries URL: https://www.accenture.com/_acnmedia/PDF-16/Accenture-World-Economic-Forum-White-Paper-January-2016-Health.pdf [accessed 2018-10-16] [WebCite Cache ID 73DJ2NLHO]
73. Stacey R. The science of complexity: an alternative perspective for strategic change processes. Strat Mgmt J 1995;16(6):477-495. [doi: 10.1002/smj.4250160606]
74. Quality Management Principles. Geneve: ISO Central Secretariat; 2012.
75. Deming W. Out of the Crisis. Cambridge, Massachustetts, London, England: MIT Press. Kindle Edition; 2000.
76. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Dec 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]

Abbreviations

Digi-PIP: Digitally supported Person-centered, Integrated, and Proactive care
eHealth: electronic health
EHR: electronic health record
ISCOPE: Integrated Systematic Care for Older People
PaJR: Patient Journey Record system
PCCP: Person-Centered Care Plan
PCMH-VA: Person-Centered Medical Home, Veterans Health Administration
PIP: Person-centered, Integrated, and Proactive
PROMs: patient-reported outcome measures
RCT: randomized controlled trial
U-PROFIT: Utrecht PROactive Frailty Intervention Trial
VA: Veterans Health Administration

Edited by CL Parra-Calderón; submitted 16.10.18; peer-reviewed by C Hernández-Quiles, P Valckenaers; comments to author 01.12.18; revised version received 01.02.19; accepted 12.02.19; published 22.04.19.

Please cite as:
Berntsen G, Strisland F, Malm-Nicolaisen K, Smaradottir B, Fensli R, Røhne M
The Evidence Base for an Ideal Care Pathway for Frail Multimorbid Elderly: Combined Scoping and Systematic Intervention Review
J Med Internet Res 2019;21(4):e12517
URL: http://www.jmir.org/2019/4/e12517/
doi:10.2196/12517
PMID:31008706

Original Paper

Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform

Jacob McPadden[1*], MD; Thomas JS Durant[2,3*], MD; Dustin R Bunch[2], PhD; Andreas Coppi[3], PhD; Nathaniel Price[4], BS; Kris Rodgerson[4]; Charles J Torre Jr[4], MSc; William Byron[4], MBA; Allen L Hsiao[1], MD; Harlan M Krumholz[3,5,6], MD; Wade L Schulz[2,3], MD, PhD

[1]Department of Pediatrics, Yale University School of Medicine, New Haven, CT, United States
[2]Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT, United States
[3]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, United States
[4]Yale New Haven Health Information Technology Services, New Haven, CT, United States
[5]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States
[6]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, United States
*these authors contributed equally

Corresponding Author:
Wade L Schulz, MD, PhD
Department of Laboratory Medicine
Yale University School of Medicine
55 Park Street PS345D
New Haven, CT, 06511
United States
Phone: 1 (203) 819 8609
Email: wade.schulz@yale.edu

Abstract

Background: Health care data are increasing in volume and complexity. Storing and analyzing these data to implement precision medicine initiatives and data-driven research has exceeded the capabilities of traditional computer systems. Modern big data platforms must be adapted to the specific demands of health care and designed for scalability and growth.

Objective: The objectives of our study were to (1) demonstrate the implementation of a data science platform built on open source technology within a large, academic health care system and (2) describe 2 computational health care applications built on such a platform.

Methods: We deployed a data science platform based on several open source technologies to support real-time, big data workloads. We developed data-acquisition workflows for Apache Storm and NiFi in Java and Python to capture patient monitoring and laboratory data for downstream analytics.

Results: Emerging data management approaches, along with open source technologies such as Hadoop, can be used to create integrated data lakes to store large, real-time datasets. This infrastructure also provides a robust analytics platform where health care and biomedical research data can be analyzed in near real time for precision medicine and computational health care use cases.

Conclusions: The implementation and use of integrated data science platforms offer organizations the opportunity to combine traditional datasets, including data from the electronic health record, with emerging big data sources, such as continuous patient monitoring and real-time laboratory results. These platforms can enable cost-effective and scalable analytics for the information that will be key to the delivery of precision medicine initiatives. Organizations that can take advantage of the technical advances found in data science platforms will have the opportunity to provide comprehensive access to health care data for computational health care and precision medicine research.

(J Med Internet Res 2019;21(4):e13043) doi:10.2196/13043

KEYWORDS

data science; monitoring, physiologic; computational health care; medical informatics computing; big data

Introduction

Background

Health care has seen massive data growth over the last several years, with some reports estimating that health care data generation increases by 48% annually [1]. In addition, it has been estimated that the intelligent use of big data within the health care sector in the United States could save over US $300 billion [2]. One particular area of medicine that relies heavily on these big data sources is precision medicine, where massive amounts of information are needed to provide precision diagnostics or therapeutics [3-5]. However, efforts to store, manage, and analyze these growing datasets have stretched the limits of traditional health care information technology systems [6].

Many definitions of big data exist, with one of the simplest being “any dataset that is too large or complex for traditional hardware or data management tools” [7]. In addition to the significant increases in volume, health care data are highly complex due to the presence of many data standards and an estimated 80% of information being unstructured [8]. These data can be problematic for traditional enterprise solutions, which rely heavily on defined data models prior to indexing, making it difficult to accommodate new data feeds or evolving data structures [9]. To support the informatics needs for the next generation of computational health research, novel approaches to data storage and analysis are necessary.

Fortunately, several applications have emerged that begin to address the key challenges in big data processing, such as distributed data storage and scalable processing capacity [10]. One example is the Hadoop platform, a set of open source tools designed specifically for these tasks [11]. The goal of these platforms is to create a central repository, called a data lake, which can store raw data in its native format for later search, retrieval, and analysis. However, researchers and clinicians in the health care sector looking to leverage modern big data architectures are faced with particular challenges in implementation and little guidance or evidence on the use of these platforms in parallel with production environments.

With the push for population-wide research initiatives in the United States such as the Cancer Moonshot [12] and Precision Medicine Initiative (now called All of Us) [3] that will rely on large, complex, interrelated data, institutions need to develop systems that can adequately scale to handle the data influx and provide sufficient capacity for analytic needs. Nevertheless, any new approaches must be attentive to the privacy and reliability requirements associated with health care data. Accordingly, we present 2 use cases that highlight the architecture and implementation of a health care data science platform that enables integrated, scalable, secure, and private health care analytics. These strategies highlight current best practices for data management, system integration, and distributed computing, while maintaining a high level of security and reliability.

Objective

We describe how an integrated data lake and analytics platform can be used to provide near real-time access to health care and biomedical research data and the ability to conduct computational health care research. We describe the implementation of such a platform, which we have named the Yale New Haven Health Baikal Data Science Platform. We highlight the data workflows and use of specialized data storage formats for 2 common health care use cases: continuous patient monitoring and real-time laboratory analytics. We chose these because they use large, real-time datasets that are difficult to store in traditional health care data warehouses.

Methods

Hardware and Operating Systems

We deployed the Hadoop platform (Apache Software Foundation) on a 30-node cluster running CentOS7 (Red Hat). No virtualization layer was used so as to minimize performance overhead. This cluster has a total of 552 processing cores, 9 TB of memory, and 714 TB of storage distributed among the nodes, with a scalable framework that can be used to add additional capacity. We deployed 5 additional nodes running CentOS7 with a distributed total of 60 cores, 320 GB of memory, and 5 TB of storage to run Elasticsearch (Elasticsearch BV). In addition to the core data storage and analysis nodes, we created 4 virtual application servers: 2 running CentOS7 and 2 running Windows Server 2012 R2 (Microsoft Corporation). We also deployed a virtual machine running CentOS7 as the Ambari management node (Apache Software Foundation) for the Hadoop cluster.

Software and System Configuration

We installed Hortonworks Data Platform version 2.6.0, a commercially supported Hadoop distribution (Hortonworks Inc), with 3 master nodes, 3 edge nodes, and 24 data nodes. We deployed Ambari with Ansible (Red Hat) playbooks and individual Hadoop applications deployed through the Ambari interface. We installed key software packages from Hortonworks Data Platform version 2.6.0, including the Hadoop Distributed File System (HDFS), Zookeeper, Yet Another Resource Negotiator (more commonly known as YARN), Kafka, Storm, and Spark, on these nodes, in high-availability mode when possible. We deployed Docker (Docker Inc) within a Swarm configuration on the 3 edge nodes. We deployed Hortonworks Data Flow (HDF) version 3.0 (Hortonworks Inc), based on the open source NiFi software (version 1.2.0.3), within a Docker container on 1 edge node.

We deployed Elasticsearch version 6.2.2 within Docker containers to 5 individual nodes. We configured 1 node as a dedicated master node, 2 nodes as master-eligible data nodes, and 2 data nodes. We deployed Kibana version 6.2.2 (Elasticsearch BV) in a Docker container to 1 Linux application server. Other software components relevant to the use cases discussed here include version 3.6 of the RabbitMQ software (Pivotal Software, Inc), the Capsule Neuron software (Qualcomm Life, Inc), and the Cloverleaf (Infor) interface and integration engine.

Physiologic monitoring data were validated by a team of nurses, respiratory therapists, engineers, and physicians simultaneously reviewing specific bedside monitor results with the real-time data feed in the data science platform. Random provider-validated data points within the electronic health record were also compared with values captured by the data science platform. Laboratory data feeds were similarly validated by physician review of randomly selected laboratory orders and results, as well as by comparison of result counts in the reference-standard clinical data warehouse and data science platform.

Compression Efficiency Assessment

We developed a Spark application in the Scala programming language to compare the storage and analytic efficiency of 3 file formats: standard comma-separated values, Avro, and data compressed with the Snappy codec (Google LLC). We extracted data for a 1-month monitoring period for performance testing. We loaded data elements and then wrote them to HDFS in each file format. We obtained the execution time for the read and write efficiency from the Spark shell interface. We repeated this process on 3 separate edge nodes and calculated the mean execution time for each measurement.

Results

Platform Architecture and Deployment

Core Components

The core software applications within the Baikal platform include features that allow for distributed data storage, message queuing, streaming data processing, distributed computation, and workflow management (Figure 1). Two key systems form the basis of the data storage platform: HDFS and Elasticsearch, a NoSQL database platform. Two message queue applications are also used within the data science platform. Kafka is used within the Hadoop environment and RabbitMQ is used on nodes outside of the Hadoop cluster. Streaming data are processed with Storm, a distributed real-time computation system, or HDF, which provides similar features but with a developer-friendly user interface. Distributed batch computation is done with the Spark framework and custom applications. Workflow management and configuration synchronization are done with the Oozie and Zookeeper applications, respectively. Finally, Docker is used for the deployment of custom applications that can be run within the data science platform.

Figure 1. Baikal platform architecture. Cluster services are monitored, deployed, and provisioned by Ambari management console. Workflow management and configuration synchronization are handled by Zookeeper and Oozie. Data storage frameworks include Hadoop Distributed File System (HDFS) and a nonrelational database: Elasticsearch. Kafka messaging queues are used for incoming data with subsequent ingest and processing handled by Storm, Sqoop, and NiFi. Analytics can be performed by Spark and Hive. Kerberos and Ranger are used to secure cluster applications. Lastly, Docker Swarm is used to deploy custom applications that can be run within the data science platform. YARN: Yet Another Resource Negotiator.

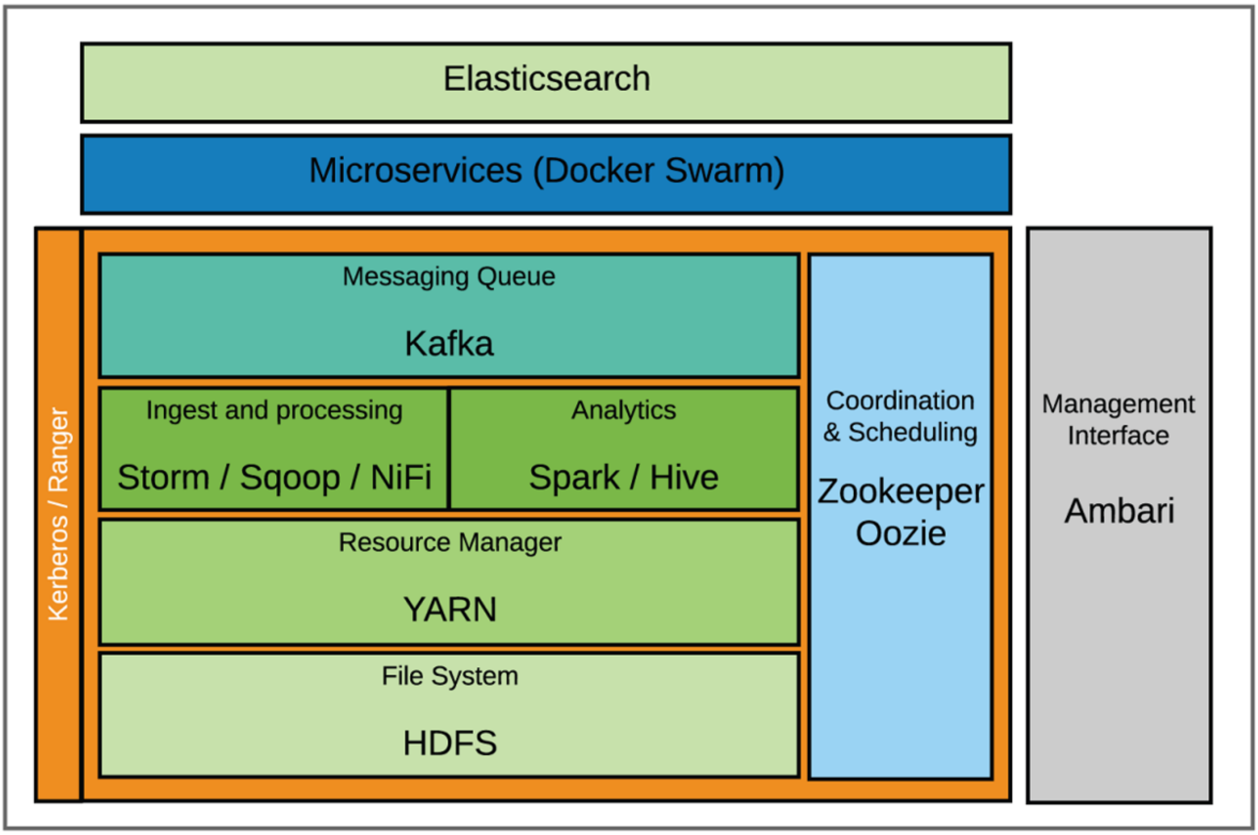


Table 1. Average storage requirements for adult and pediatric patient monitoring and ventilator monitoring. Signal frequency and storage size are the metrics for a complete 24-hour per-bed monitoring period averaged from 3 independent samples.

Source	Signal frequency, mean (SD)	Storage size (MB), mean (SD)	Estimated annual storage (GB)
Adult monitor	291,252 (84,568)	17.1 (5.0)	6.2
Pediatric monitor	223,387 (29,543)	12.7 (1.8)	4.6
Adult ventilator	3,504,162 (236,672)	231.5 (30.6)	84.5

Security

Many big data platforms, including Hadoop, have limited security features enabled by default [13,14]. For example, no authentication is required to access Web service interfaces by default in either Hadoop or Elasticsearch [14,15]. This lack of default security has led to several data breaches over the last several years [16]. Fortunately, these platforms do allow for configuration of a robust security system with the use of Kerberos, Ranger, and Shield [13,17]. Within the Baikal platform, we deployed a dedicated Kerberos realm for authentication into the cluster. We deployed Ranger to allow for permission-based authorization to resources in the cluster at both the application and data layers.

Electron: A Framework for Physiologic Signal Monitoring and Analysis

Continuous monitoring of patient vital signs has been standard practice in intensive care units and emergency departments. However, these data are rarely kept longer than a few days due to the storage and technical requirements for such large datasets with limited impact for clinical use; however, they may have significant value for future investigation. To support investigators who require access to these physiologic signals, we used the Baikal platform to create Electron, a framework to store and analyze longitudinal physiologic monitoring data. The code for this platform is available within a GitHub repository [18].

Data Characteristics

Many bedside patient monitors and ventilators are able to transmit their settings and recordings to a central application at specific intervals. Within our institution, these signals are sent at 1- to 5-second intervals, depending on the device, data element, and value. These data elements include active data channels, device and patient metadata, and more intermittent data elements, such as noninvasive blood pressure and alarms. In total, data can be transmitted for up to 892 active and metadata channels for bedside monitors and 45 channels for ventilators. Individual message size varies based on the number of metadata elements, the device being used for monitoring, and frequency of intermittent events. To determine the data storage requirements for such a platform, we collected data for 3 randomly selected adult and 3 pediatric patients in the intensive care unit for a 24-hour time frame. A single adult patient in the intensive care unit generated approximately 17.1 MB of data per 24-hour time frame (Table 1).

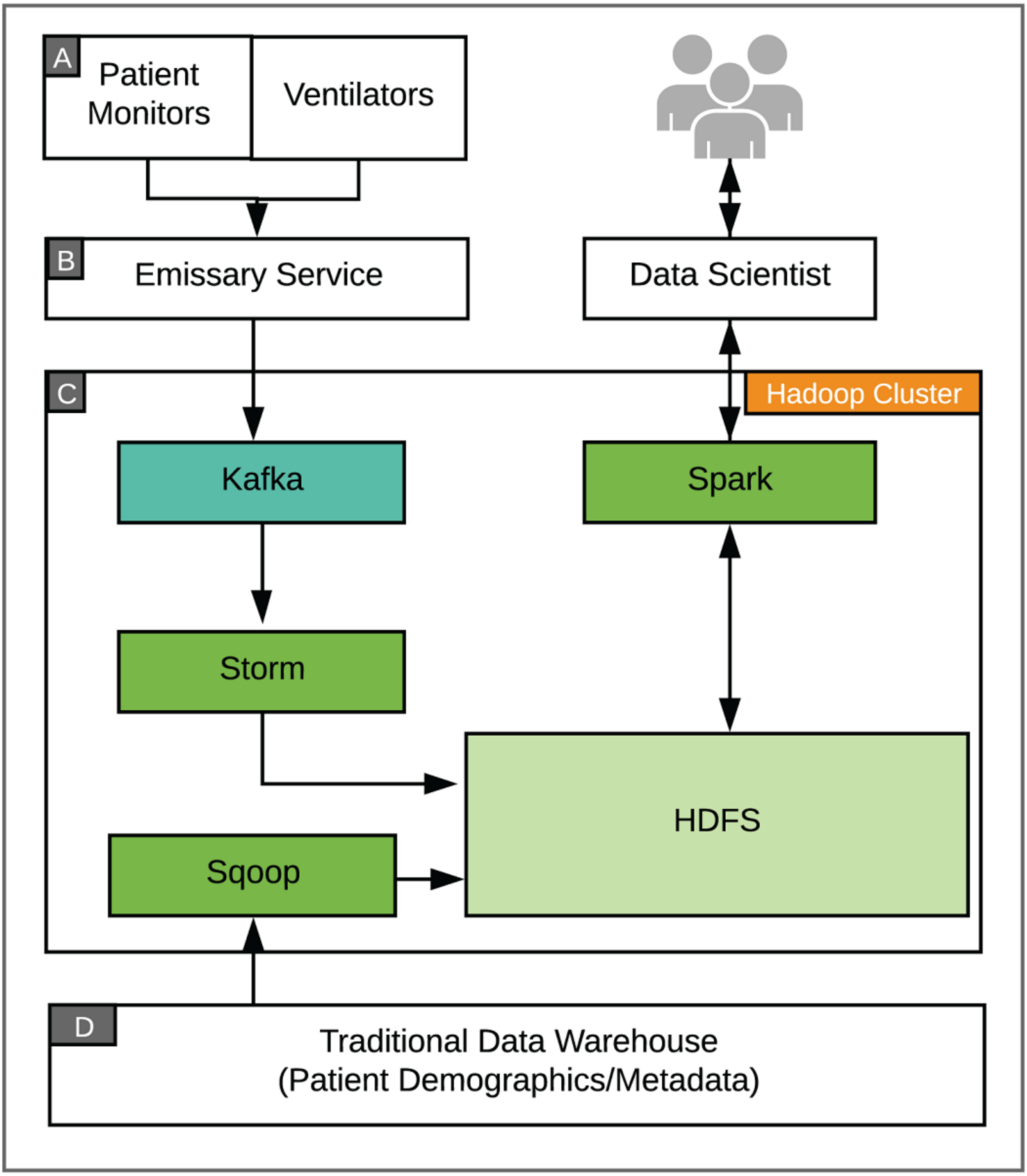
Similarly, a 24-hour monitoring period for pediatric patients averaged approximately 12.7 MB in the same time frame. We similarly assessed the data volume generated by ventilators in our health system, which produced approximately 231.5 MB of data per ventilator per day. When we assessed monitoring data from a 1-month period, we collected over 6 TB of raw data from 11 units and a total of 225 beds, often reaching rates of over 400 messages/s. These units were diverse and included intensive care, surgical, emergency department, and short-stay beds.

Electron Framework Architecture

The platform to acquire, store, and analyze the continuous monitoring data consists of 4 key features: data ingestion, data processing and denormalization, compressed storage, and distributed analytics (Figure 2). Our physiologic monitoring infrastructure consists of attached patient monitoring devices that send signals to vendor-supported integration servers (Figure 2, box A). Data are then transmitted as Health-Level 7 (HL7) messages streamed via a Transmission Control Protocol/Internet Protocol connection to an emissary service that we deployed to accept the incoming message stream and perform the initial conversion of HL7 messages into a custom JSON format (Figure 2, box B). Date and time information is converted to Coordinated Universal Time (UTC), while all other data are left in their original format. Once processed, messages are forwarded to a secured Kafka message queue, which allows the platform to buffer messages during downstream processor downtime or when under heavy load. The JSON document contains key elements for downstream processing, as well as a copy of the original HL7 message to allow for future reprocessing, if needed (Textbox 1).

The decision to store the original data is often considered a best practice but has the disadvantage of increasing the storage requirements for the dataset. While we opted to store the raw data on the platform, as the HL7 message contains additional information that may be needed for future studies, the decision to maintain this information can be made for each use case dependent on the data storage costs and estimated future utility of the information.

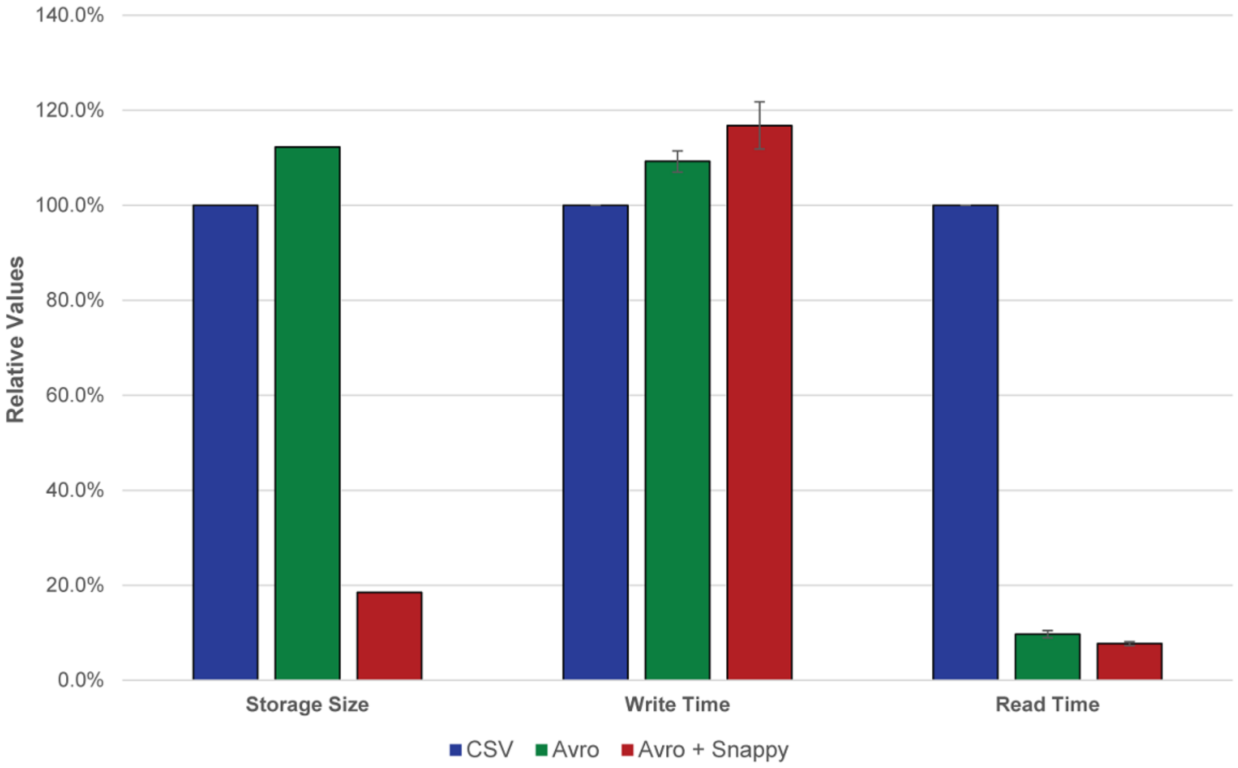
Figure 2. System architecture for continuous patient monitoring. Multiple, increasing sources of clinical data (A) acquire and transmit the data to aggregation servers, which then forward Health-Level 7 (HL7) messages to an emissary service (B), where data are normalized and securely forwarded in standardized JSON format to the Baikal system (C) for denormalization, processing, and storage in the Hadoop Distributed File System (HDFS). Traditional historic databases (D) are individually prepared for ingestion in the Baikal system and storage in HDFS. The resulting data lake allows for integrated, distributed analytics by end users.



Textbox 1. JSON data model for physiologic data.

```
{
"msh_ts":"long",
"alarm_ts":"long",
"source":"string",
"unit":"string",
"text":"string",
"channel":"string",
"hl7":"string"
}
```

Figure 3. Comparison of storage and read/write efficiency. Avro increases storage space and write time modestly while significantly reducing read time. The addition of Snappy compression increases write time minimally, while significantly decreasing storage space and maintaining minimal read time. The resulting combination optimizes for single archival write with multiple read usage. CSV: comma-separated values. Error bars represent standard error.



Storage Formats and Data Processing

While storage costs continue to decline, the cost of long-term data storage for large datasets remains burdensome. Specialized data formats and compression can improve the density of data storage but often come at the cost of increased overhead for read and write throughput. Fortunately, the frequency of access to historic data typically decreases over time, which means that slower data access methods would have less impact on overall analytic capacity. Other work has compared the storage and access efficiency for many big data technologies [19]. For this use case, we predicted that the Avro data format with Snappy compression would have an appropriate balance of storage and access efficiency.

Avro is a semi-structured data serialization format designed for big data storage. In addition to the semi-structured nature of the Avro format, the files are also splittable, which means that the Hadoop platform can separate the file into individual sections, which increases the processing efficiency during data analysis [20]. To assess the impact of the Avro format and Snappy compression, we assessed the storage and access efficiency of monitoring data from several different variables over a 30-day period in comma-separated raw text, Avro, and Snappy-compressed Avro formats. Data were filtered and the length of time needed to write and read data from 3 independent nodes in the cluster was recorded. Compared with raw text, Avro-formatted files required approximately 12% more storage space on disk but showed significantly faster data retrieval time (Figure 3). The use of Snappy compression showed significant savings in storage requirements, with an average reduction in file size of 80.5% compared with raw text files. Also of note was the large reduction in time needed to access data stored in Snappy-compressed Avro files.

In addition to the large volume, the high velocity of these data required a high-throughput data processing pipeline to convert and store the data efficiently. To achieve this, we developed a custom application built on the Storm platform that allowed for distributed, high-throughput processing. Within the Storm topology, monitor signals were denormalized, converted to the Avro format, compressed with the Snappy codec, and stored in HDFS to allow for future analysis. A separate copy of the data containing the original HL7 message was also stored through a separate Storm bolt in case reprocessing of the raw data became necessary.

Analytics

Much like the particular challenges for the acquisition and storage of big data, specialized needs for the analysis of these datasets also exist. While the raw data are of use for many research and clinical projects, derived variables and predictive analytics are often of interest. For example, computationally derived features, such as R-R intervals [21], indices of multiple vital signs [22], and temporal relations between vital signs have all shown promise as predictive variables [23]. However, generating these features is often computationally intensive when performed at scale on entire patient populations.

Traditional analytic methods and tools are often unable to scale to meet the needs of these analyses. Even in cases where parallelized computation can be used, the resources necessary to develop and validate these custom applications is often prohibitive. To make parallelized computation more accessible, solutions such as MapReduce [11] and Dryad [24] have been created, which provide frameworks that manage the complexity of parallelization. However, these solutions still require significant technical expertise to develop applications that can be deployed to production environments. Within the Baikal platform, we enabled Spark as the primary data analysis tool for batch analysis. Spark is a general data processing framework that can be used to write applications in several common languages, including Java, Scala, Python, and R. A key advantage of this framework is the ability to maintain data for MapReduce operations in memory, rather than needing to read and write each intermediate step to disk. This has been shown to improve the speed of big data processing significantly [25,26]. We developed several Spark applications that can be used by data analysts to generate features from the physiologic data, such as alarm events, and extract subsets of information for downstream processing, which are available within the GitHub repository.

While the physiologic data are captured during routine clinical care, a major goal of this dataset is for use in biomedical research. Because of the sensitive nature and regulatory oversight of human subjects research, appropriate approvals, such as institutional review board approval and patient consent, are needed prior to the analysis of data within the system. In addition, our clinical data warehouse (Figure 2, box D) includes a field that indicates whether patients have explicitly opted out of research and is used to filter their data from analysis that is classified as human subjects research.

Nucleus: A Platform for Real-Time Laboratory Business Intelligence and Data Visualization

In addition to novel data sources such as continuous patient monitoring, data science platforms can also offer new approaches for the analysis of more traditional health care datasets. Examples include real-time data analysis, predictive analytics, and interactive visualizations. In the era of cost reduction and an increasing demand for clinical laboratory services, laboratorians are facing expectations to optimize laboratory efficiency for the sake of clinical workflows and improve test utilization without compromising quality and safety. Therefore, the clinical laboratory has a particular need for real-time business intelligence to improve testing efficiency and patient safety [27]. To achieve this, we created a data science platform with business intelligence dashboards to monitor testing within our institution's clinical laboratory [28].

Data Characteristics

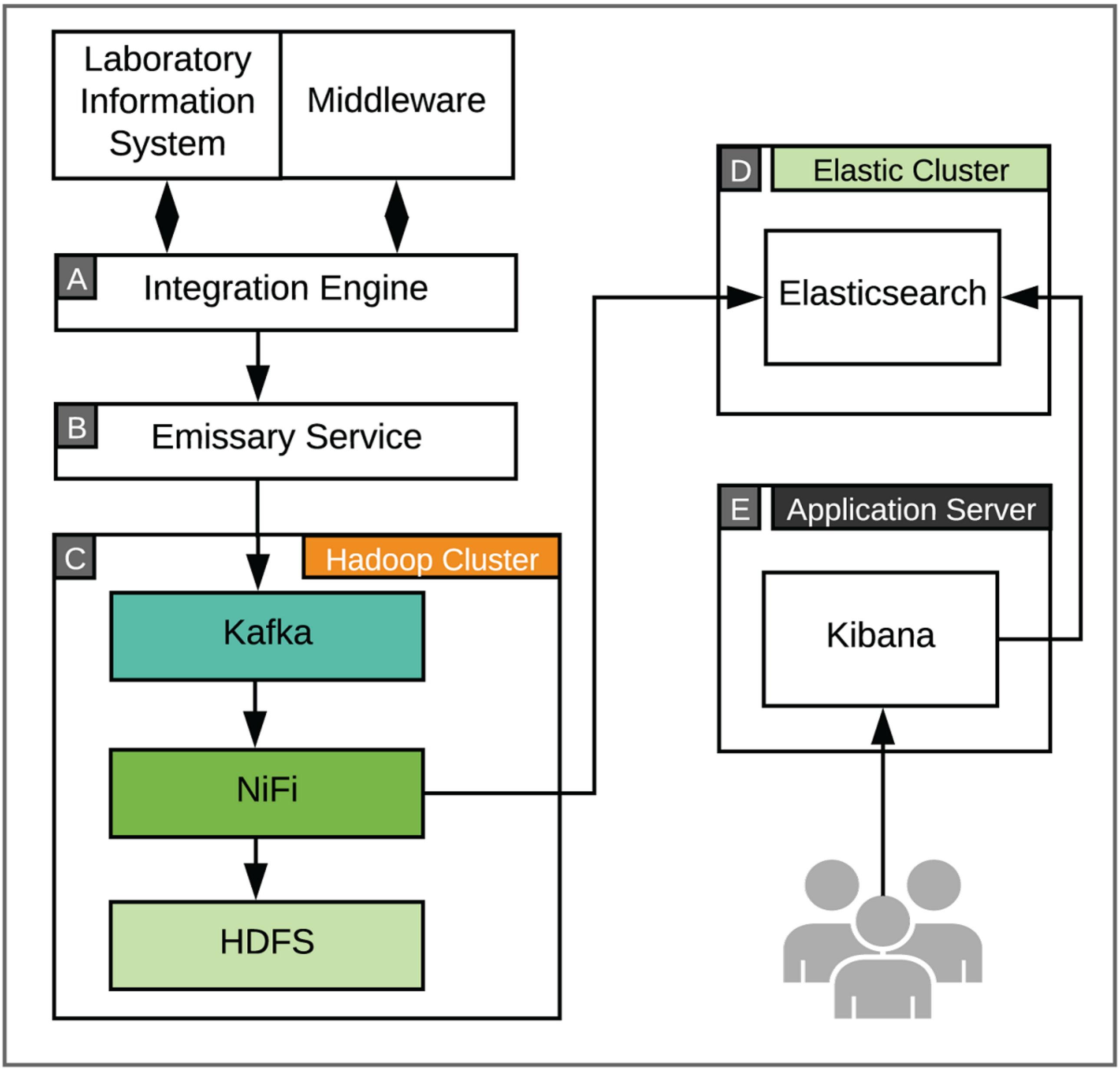
Laboratory orders and results are often routed through multiple systems as they transit between the electronic health record and laboratory instrumentation. This typically includes message integration services and middleware platforms that manage the flow of data between systems created by a number of different vendors. Within our health care system, approximately 40 million individual results are generated annually from 6 hospitals, 26 satellite locations, and approximately 220 laboratory instruments. A principal challenge for these data is to provide real-time access and visualizations to end users who need actionable insights from these disparate systems. Because of these unique needs, many downstream architectural decisions varied from the continuous monitoring application described.

Nucleus Platform Architecture

The initial acquisition of data for this stream is similar to that of the continuous monitoring workflow (Figure 4). Briefly, we deployed an emissary service to receive an HL7 stream of clinical observations and results messages from the Cloverleaf integration engine. Each HL7 message was validated and mapped to a JSON document by the emissary service, then forwarded to a secured Kafka message queue. The custom JSON messages contain key parameters that can be used to index and parse results during batch analysis (Textbox 2).

Because of the slower message velocity, this data stream was easily processed with the NiFi/HDF software, which is designed for real-time data processing. We created custom Python scripts to process and denormalize the incoming data stream. Each order and result message was then written to HDFS for permanent storage and batch analysis and also routed to Elasticsearch for real-time analysis and visualization. Additional features that provide key indicators of laboratory efficiency were generated in real time from the HL7 messages with custom Python scripts that are executed within the NiFi workflow. These quality indicators are stored within Elasticsearch and can be used to visualize turnaround time for laboratory results, outstanding orders, and order volumes by patient or laboratory location.

Figure 4. System architecture for laboratory data monitoring. Health-Level 7 (HL7) observations and results messages generated by laboratory information system and laboratory middleware systems are received by the clinical integration engine Cloverleaf (A). HL7 messages are received and validated by a custom emissary service (B) and mapped to JSON documents, which are submitted to a Kafka message queue for downstream processing (C). Custom Python (version 2.7) scripts are executed in NiFi to denormalize messages and calculate quality improvement metrics. Raw HL7 messages are stored in a Hadoop Distributed File System (HDFS). Processed messages and quality improvement metrics are routed to Elasticsearch (D) for real-time analysis and Kibana (E) for visualization.



Textbox 2. JSON data model for laboratory data.

```
{
“msh_ts”:"long",
“pt_mrn”:“string”,
“order_id”:“string”,
“lab_type_code”:“string”,
“order_ts”:"long",
“hl7”:“string”
}
```

Discussion

Principal Findings

Health care information is inherently complex and often has an evolving data structure; much of the data is not stored in the electronic health record. Because of this, novel approaches to data management are needed to integrate the many sources of health care data. In addition, novel approaches to data analysis such as machine learning require significant computational resources for timely analysis. As the use of big data in health care continues to increase, the implementation of robust technical solutions to manage and analyze the data will be important to the success of biomedical big data research [3].

In this paper, we have presented the successful implementation of a data science platform along with 2 domain-specific applications deployed within this platform. These applications focused on the storage of high volume, real-time datasets that challenge traditional data warehousing strategies due to their volume and velocity. We have also presented the hardware and architectural approaches used to manage these data. While individual components of the platform used here are described in the nonmedical literature, this platform combines available technologies to meet the known challenges of big data with needs specific to health care, including the security and privacy needs of personal health information.

Often, a single technical solution is unable to address all concerns or needs for a robust data science environment. For example, Hadoop has traditionally been used as a platform for big data storage and batch analysis but had fewer tools available for streaming data and real-time analytics. Because of this, we integrated components designed specifically for the management and visualization of real-time data. This integration allows us to provide efficient batch analytics, as well as real-time visualizations, which would be challenging if only a single tool or platform were used. It should be noted, however, that the applications described here are rapidly evolving and significant strides have been made to expand the features of each component, which may add redundancy between applications in the future.

Data science platforms such as Hadoop offer many individual components to address key requirements for data replication, availability, and security at each stage of the data life cycle, from acquisition to analysis. Fully implementing each of these utilities can make data science pipelines complex, but the use of service-oriented architectures affords the ability to update individual applications, scale services, and reuse individual components in multiple workflows. Because of these rapid developments and the diversity of data, careful testing should be done during the implementation of data science workflows to determine the storage and compute the capacity required for long-term management of the data being obtained. Similarly, careful attention should be paid to the implementation of built-in security features to ensure that data are not accidentally made available to unauthorized users [13].

Limitations

While data science platforms offer significant potential for the rapid analysis of big data, several limitations exist. In particular, the complexity of these platforms often requires substantial technical expertise to use them to their full potential. Multiple software applications are often needed to implement an entire workflow, particularly within the Hadoop environment. While each Hadoop component often provides significant advantages from developing new applications, personnel with expertise are needed to implement these technologies effectively. While many attempts have been made to make the environment fluent with other tools, such as Python, SAS, and R, seamless integration with these tools remains difficult, particularly in secured environments.

Massive resources have been dedicated to big data and data science in other industries; however, the return on investment has not always been realized. Therefore, the ultimate success of these platforms for computational health research will depend on the ability of the biomedical research community to apply big data to translational and clinical research. Successful application of these technologies with applications that can provide actionable insights from real-world data has the potential to deliver precision medicine at the point of care, but additional studies will be needed to fully assess the impact of these systems on health care delivery and clinical outcomes.

Conclusion

The paucity of literature describing implementation experiences leaves those interested in developing big data environments largely unguided, particularly within the health care sector, which has unique data and regulatory requirements. Careful attention to the architecture used to create these data science environments will provide an important foundation for future studies that create value from big data sources. As the volume and velocity of health care data continue to increase, additional analyses on the management of these data will be required to ensure that the highest-quality data are made available to efficient analytic systems.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to acknowledge Steven Shane and Richard Hurley from Yale New Haven Health Information Technology Services, New Haven, CT, USA, for their assistance with platform implementation and maintenance. We would also like to thank Patrick Young and Guannan Gong for their review of the manuscript.

Conflicts of Interest

HMK was a recipient of a research grant, through Yale, from Medtronic and the US Food and Drug Administration to develop methods for postmarket surveillance of medical devices; is a recipient of research agreements with Medtronic and Johnson & Johnson (Janssen), through Yale, to develop methods of clinical trial data sharing; works under contract with the US Centers for Medicare & Medicaid Services to develop and maintain performance measures that are publicly reported; chairs a Cardiac Scientific Advisory Board for UnitedHealth Group Inc; is a participant and participant representative of the IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element Science, Inc, and the Physician Advisory Board for Aetna Inc; and is the founder of Hugo, a personal health information platform. WLS is a consultant for Hugo, a personal health information platform.

References

1. EMC. The digital universe driving data growth in healthcare. Hopkinton, MA: Dell Inc; 2014. URL: https://www.emc.com/analyst-report/digital-universe-healthcare-vertical-report-ar.pdf [accessed 2018-10-03] [WebCite Cache ID 72tbbQNwk]
2. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: the next frontier for innovation, competition, and productivity. New York, NY: McKinsey Global Institute; 2011 Jun. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx [accessed 2019-02-04] [WebCite Cache ID 75wC2TnLO]
3. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
4. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. N Engl J Med 2015 Jun 04;372(23):2229-2234. [doi: 10.1056/NEJMsb1503104] [Medline: 26014593]
5. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. Proteomics 2016 Mar;16(5):741-758. [doi: 10.1002/pmic.201500396] [Medline: 26677817]
6. Jordan L. The problem with big data in translational medicine. A review of where we've been and the possibilities ahead. Appl Transl Genom 2015 Sep;6:3-6 [FREE Full text] [doi: 10.1016/j.atg.2015.07.005] [Medline: 27054070]
7. Sun J, Reddy CK. Big data analytics for healthcare. 2013 Presented at: 19th ACM SIGKDD international conference on Knowledge discovery and data mining; Aug 11-14, 2013; Chicago, IL, USA. [doi: 10.1145/2487575.2506178]
8. IBM. Data-Driven Healthcare Organizations Use Big Data Analytics for Big Gains. Armonk, NY: IBM Corporation; 2013.
9. Fang H. Managing data lakes in big data era: what's a data lake and why has it became popular in data management ecosystem. 2015 Presented at: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER); Jun 8-12, 2015; Shenyang, China. [doi: 10.1109/CYBER.2015.7288049]
10. Patel AB, Birla M, Nair U. Addressing big data problem using Hadoop and Map Reduce. 2012 Presented at: 2012 Nirma University International Conference on Engineering (NUiCONE); Dec 6-8, 2012; Ahmedabad, India. [doi: 10.1109/NUICONE.2012.6493198]
11. Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun ACM 2008;51(1):107-113.
12. Fact sheet: investing in the national Cancer Moonshot. Washington, DC: The White House, Office of the Press Secretary; 2016 Feb 01. URL: https://obamawhitehouse.archives.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot [accessed 2019-02-04] [WebCite Cache ID 75wCltSOe]
13. Spivey B, Echeverria J. Hadoop Security: Protecting Your Big Data Platform. Sebastopol, CA: O'Reilly Media, Inc; 2015.
14. Sharma PP, Navdeti CP. Securing big data Hadoop: a review of security issues, threats and solution. Int J Comput Sci Inf Technol 2014;5(2):2126-2131.
15. Inukollu VN, Arsi S, Ravuri SR. Security issues associated with big data in cloud computing. Int J Netw Secur Appl 2014;6(3):45.
16. Matherly J. Shodan blog. 2017 May 31. The HDFS juggernaut URL: https://blog.shodan.io/the-hdfs-juggernaut/ [accessed 2018-10-03] [WebCite Cache ID 72tbk0pY4]
17. Derbeko P, Dolev S, Gudes E, Sharma S. Security and privacy aspects in MapReduce on clouds: a survey. Comput Sci Rev 2016;20:1-28.
18. Schulz W. Electron. San Francisco, CA: GitHub, Inc; 2018 Aug 12. URL: https://github.com/ComputationalHealth/electron [accessed 2019-02-05] [WebCite Cache ID 75xwVN7o9]
19. Grover M, Malaska T, Seidman J, Shapira G. Hadoop Application Architectures. Sebastopol, CA: O'Reilly Media, Inc; 2015.
20. Apache Avro 1.8.1 specification. 2016. URL: https://avro.apache.org/docs/1.8.1/spec.html [accessed 2018-10-03] [WebCite Cache ID 72tbp0DbK]
21. Griffin MP, Moorman JR. Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. Pediatrics 2001 Jan;107(1):97-104. [Medline: 11134441]

22. Parimi N, Hu PF, Mackenzie CF, Yang S, Bartlett ST, Scalea TM, et al. Automated continuous vital signs predict use of uncrossed matched blood and massive transfusion following trauma. J Trauma Acute Care Surg 2016 Dec;80(6):897-906. [doi: 10.1097/TA.0000000000001047] [Medline: 27027555]
23. Bashan A, Bartsch RP, Kantelhardt JW, Havlin S, Ivanov PC. Network physiology reveals relations between network topology and physiological function. Nat Commun 2012 Feb 28;3:702 [FREE Full text] [doi: 10.1038/ncomms1705] [Medline: 22426223]
24. Isard M, Budiu M, Yu Y, Birrell A, Fetterly D. Dryad: distributed data-parallel programs from sequential building blocks. 2007 Presented at: 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems; Mar 21-23, 2007; Lisbon, Portugal.
25. Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Stoica I. Spark: cluster computing with working sets. 2010 Presented at: HotCloud ’10. 2nd USENIX conference on Hot topics in cloud computing; Jun 22-25, 2010; Boston, MA, USA.
26. Zaharia M, Chowdhury M, Das T, Dave A, Ma J, McCauley M, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. 2012 Presented at: 9th USENIX Symposium on Networked Systems Design and Implementation; Apr 25-27, 2012; San Jose, CA, USA.
27. Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem 2002 May;48(5):691-698. [Medline: 11978595]
28. Schulz W. Nucleus. San Francisco, CA: GitHub, Inc; 2018 Aug 13. URL: https://github.com/ComputationalHealth/nucleus [accessed 2019-02-05] [WebCite Cache ID 75xwZ0Qo8]

Abbreviations

HDF: Hortonworks Data Flow
HDFS: Hadoop Distributed File System
HL7: Health-Level 7
YARN: Yet Another Resource Negotiator

Edited by G Eysenbach; submitted 06.12.18; peer-reviewed by J Rudolf, A Benis; comments to author 26.12.18; revised version received 09.01.19; accepted 13.01.19; published 09.04.19.

Please cite as:
McPadden J, Durant TJS, Bunch DR, Coppi A, Price N, Rodgerson K, Torre Jr CJ, Byron W, Hsiao AL, Krumholz HM, Schulz WL
Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform
J Med Internet Res 2019;21(4):e13043
URL: https://www.jmir.org/2019/4/e13043/
doi:10.2196/13043
PMID:30964441

Review

Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Pawel Posadzki[1], PhD; Malgorzata M Bala[2], MD, PhD; Bhone Myint Kyaw[3], PhD, MBBS; Monika Semwal[1], PhD; Ushashree Divakar[1], MSc; Magdalena Koperny[4], PhD; Agnieszka Sliwka[5], PhD; Josip Car[1], MD, PhD, FRCP, FFPH

[1]Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland
[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[4]Province Sanitary-Epidemiological Station of Lesser Poland, Public Health and Health Promotion Department, Krakow, Poland
[5]Institute of Physiotherapy, Faculty of Health Sciences, Jagiellonian University Medical College, Krakow, Poland

Corresponding Author:
Josip Car, MD, PhD, FRCP, FFPH
Centre for Population Health Sciences (CePHaS)
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
Singapore
Singapore
Phone: 65 6904 7005
Email: josip.car@ntu.edu.sg

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/6/e20316

Abstract

Background: The shortage and disproportionate distribution of health care workers worldwide is further aggravated by the inadequacy of training programs, difficulties in implementing conventional curricula, deficiencies in learning infrastructure, or a lack of essential equipment. Offline digital education has the potential to improve the quality of health professions education.

Objective: The primary objective of this systematic review was to evaluate the effectiveness of offline digital education compared with various controls in improving learners' knowledge, skills, attitudes, satisfaction, and patient-related outcomes. The secondary objectives were (1) to assess the cost-effectiveness of the interventions and (2) to assess adverse effects of the interventions on patients and learners.

Methods: We searched 7 electronic databases and 2 trial registries for randomized controlled trials published between January 1990 and August 2017. We used Cochrane systematic review methods.

Results: A total of 27 trials involving 4618 individuals were included in this systematic review. Meta-analyses found that compared with no intervention, offline digital education (CD-ROM) may increase knowledge in nurses (standardized mean difference [SMD]=1.88; 95% CI 1.14 to 2.62; participants=300; studies=3; I^2=80%; low certainty evidence). A meta-analysis of 2 studies found that compared with no intervention, the effects of offline digital education (computer-assisted training [CAT]) on nurses and physical therapists' knowledge were uncertain (SMD 0.55; 95% CI –0.39 to 1.50; participants=64; I^2=71%; very low certainty evidence). A meta-analysis of 2 studies found that compared with traditional learning, a PowerPoint presentation may improve the knowledge of patient care personnel and pharmacists (SMD 0.76; 95% CI 0.29 to 1.23; participants=167; I^2=54%; low certainty evidence). A meta-analysis of 4 studies found that compared with traditional training, the effects of computer-assisted training on skills in community (mental health) therapists, nurses, and pharmacists were uncertain (SMD 0.45; 95% CI –0.35 to 1.25; participants=229; I^2=88%; very low certainty evidence). A meta-analysis of 4 studies found that compared with traditional training, offline digital education may have little effect or no difference on satisfaction scores in nurses and mental health therapists (SMD –0.07; 95% CI –0.42 to 0.28, participants=232; I^2=41%; low certainty evidence). A total of 2 studies found that offline digital education may have little or no effect on patient-centered outcomes when compared with blended learning. For skills and attitudes, the results were mixed and inconclusive. None of the studies reported adverse or unintended effects of the interventions. Only 1 study reported costs of interventions. The risk of bias was predominantly unclear and the certainty of the evidence ranged from low to very low.

Conclusions: There is some evidence to support the effectiveness of offline digital education in improving learners' knowledge and insufficient quality and quantity evidence for the other outcomes. Future high-quality studies are needed to increase generalizability and inform use of this modality of education.

(J Med Internet Res 2019;21(4):e12968) doi:10.2196/12968

KEYWORDS

randomized controlled trial; systematic review; medical education

Introduction

Background

There is no health care system without health professionals. The health outcomes of people rely on well-educated nurses, pharmacists, dentists, and other allied health professionals [1]. Unfortunately, these professionals are in short supply and high demand [2,3]. Almost 1 billion people are negatively affected by the lack of access to adequately trained health professionals, suffering ill-health or dying [4,5]. In many low- and middle-income countries (LMICs), this situation is further aggravated by the difficulties in implementing traditional learning programs; deficiencies in health care systems and infrastructure; and lack of essential supplies, poor management, corruption, or low remuneration [6].

Digital education also known as e-learning is an umbrella term encompassing a broad spectrum of educational interventions characterized by their tools, technological contents, learning objectives or outcomes, pedagogical approaches, and delivery settings, which includes, but is not limited to, online and offline computer-based digital education, massive open online courses (MOOCs), mobile learning (mLearning), serious gaming and gamification, digital psychomotor skill trainers, virtual reality, or virtual patient scenarios [7]. Digital education aims to improve the quality of teaching by facilitating access to resources and services, as well as remote exchange of information and peer-to-peer collaboration [8]; it is also being increasingly recognized as one of the key strategic platforms to build strong education and training systems for health professionals worldwide [9]. The United Nations and the World Health Organization consider digital education as an effective means of addressing the educational needs among health professionals, especially in LMICs.

This review focused on offline digital education. This refers to the use of personal computers or laptops to assist in delivering stand-alone multimedia materials without the need for the internet or local area network connections [10]. The educational content can be delivered via videoconferences, emails, and audio-visual learning materials kept in either magnetic storage, for example, floppy disks, or optical storage, for example, CD-ROM, digital versatile disk, flash memory, multimedia cards, external hard disks, or downloaded from a networked connection, as long as the learning activities do not rely on this connection [11].

There are several potential benefits of offline digital education such as unrestrained knowledge transfer, enriched accessibility, and significance of health professions education [12]. Further benefits include flexibility and adaptability of educational content [13], so that learners can absorb curricula at a convenient pace, place, and time [14]. The interventions can also be used to deliver an interactive, an associative, and a perceptual learning experience by combining text, images, audio, and video via combined visual, auditory, and spatial components, further improving health professionals' learning outcomes [15,16]. By doing so, offline digital education can potentially stimulate neurocognitive development (memory, thinking, and attention) by enhancing changes in the efficiency of chemical synaptic transmission between neurons, increasing specific neuronal connections and creating new patterns of neuronal connectivity and generating new neurons [17]. Finally, health professionals better equipped with knowledge, skills, or professional attitudes as a result of offline digital education might improve the quality of health care services provision, as well as the patient-centered and public health outcomes, and reduce the costs of health care.

Objectives

This systematic review was one of a series of reviews evaluating the scope for implementation and the potential impact of a wide range of digital health education interventions for postregistration and preregistration health professionals. The objective of this systematic review was to evaluate the effectiveness of offline digital education compared with various controls in improving learners' knowledge, skills, attitudes, satisfaction, and patient-centered outcomes.

Methods

At the time of conducting and reporting the review, we used and adhered to the systematic review methods as recommended by the Cochrane Collaboration [18]. For a detailed description of the methodology, please refer to the study by Car et al [7].

Search Strategy and Data Sources

We searched the following databases (from January 1990 to August 2017): MEDLINE (via Ovid), Excerpta Medica dataBASE (via Elsevier), Web of Science, Educational Resource Information Center (via Ovid), Cochrane Central Register of

Controlled Trials, The Cochrane Library, PsycINFO (via Ovid), and the Cumulative Index to Nursing and Allied Health Literature (via EBSCO). The search strategy for MEDLINE is presented in Multimedia Appendix 1. We searched for papers in English but considered eligible studies in any language. We also searched 2 trial registries (EU Clinical Trials Register and ClinicalTrials.gov), screened reference lists of all included studies and pertinent systematic reviews, and contacted the relevant investigators for further information.

Eligibility Criteria

Only randomized controlled trials (RCTs) and cluster RCTs (cRCTs) of postregistration health professionals except medical doctors—as they were covered in a separate review [19]—using either stand-alone or blended offline digital education with any type of controls (active or inactive) measuring knowledge, skills, attitudes, satisfaction, and patient-centered outcomes (as primary outcomes) as well as adverse effects or costs (as secondary outcomes) were eligible for inclusion in this review.

We excluded crossover trials, stepped wedge design, interrupted time series, controlled before and after studies, and studies of doctors (including medical diagnostics and treatment technologies) or medical students. Participants were not excluded on the basis of sociodemographic characteristics such as age, gender, ethnicity, or any other related characteristics.

Data Selection, Extraction, and Management

The search results from the different electronic databases were combined in a single EndNote (X8.2) library, and duplicate records of the same reports were removed. In total, 2 reviewers independently screened titles and abstracts to identify studies that potentially meet the inclusion criteria. The full text versions of these articles were retrieved and read in full. Finally, 2 review authors independently assessed articles against the eligibility criteria, and 2 reviewers independently extracted the data for each of the included studies using a structured data extraction form and the Covidence Web-based software (Veritas Health Innovation, Melbourne, Australia). We extracted all relevant data on the characteristics of participants, intervention, comparator group, and outcome measures. For continuous data, we reported means and SDs and odds ratios (ORs) and its 95% CIs for dichotomous data. For studies with multiple arms, we compared the relevant intervention arm to the least active control arm, so that double counting of data does not occur. Any disagreements were resolved through discussion between the 2 authors and if no consensus was reached, a third author acted as an arbiter.

Assessment of Risk of Bias

In total, 2 reviewers independently assessed the risk of bias of the included studies using the Cochrane Collaboration’s *Risk of Bias* tool [18]. Studies were assessed for risk of bias in the following domains: random sequence generation; allocation concealment; blinding of participants or personnel; blinding of outcome assessment; completeness of outcome data (attrition bias); selective outcome reporting (reporting bias); validity and reliability of outcome measures; baseline comparability; and consistency in intervention delivery. For cRCTs, we also assessed and reported the risk of bias associated with an additional domain: selective recruitment of cluster participants. Judgments concerning the risk of bias for each study fell under 3 categories: high, low, or unclear risk of bias.

Data Synthesis

Data were synthesized using Review Manager version 5.3. In cases where studies were homogeneous enough (in terms of their population interventions, comparator groups, outcomes, and study designs) to make meaningful conclusions, we pooled them together in a meta-analysis using a random-effects model and presented results as standardized mean difference (SMD). We assessed heterogeneity through a visual inspection of the overlap of forest plots and by calculating the chi-square tests and I^2 inconsistency statistics [18].

Summary of Findings Tables

We prepared the Summary of Findings (SoF) tables to present the results for each of the primary outcomes. We converted results into absolute effects when possible and provided a source and rationale for each assumed risk cited in the table(s) when presented. A total of 2 authors (PP and MS) independently rated the overall quality of evidence as implemented and described in the GRADEprofiler (GRADEproGDT Web-based version) and Chapter 11 of the *Cochrane Handbook for Systematic Reviews of Interventions* [20]. We considered the following criteria to assess the quality of evidence: limitations of studies (risk of bias), inconsistency of results, indirectness of the evidence, imprecision and publication bias, and downgraded the quality where appropriate. This was done for all primary outcomes reported in the review.

Results

Our searches yielded a total of 30,532 citations; and 27 studies with 4,618 participants are included in Figure 1. For characteristics of excluded studies, please refer to Multimedia Appendix 2.

Figure 1. Study flow diagram. RCT: randomized controlled trial.

Included Studies

Details of each trial are presented in Table 1 or Multimedia Appendix 3; a summary is given below. The included trials were published between 1991 and 2016, and originated from Brazil 3.7% (n=1), Hong Kong 3.7% (n=1), Iran 7.4% (n=2), Korea 3.7% (n=1), the Netherlands 11.1% (n=3), Norway 3.7% (n=1), Taiwan 11.1% (n=3), Turkey 3.7% (n=1), the United Kingdom 7.4% (n=2), and the United States 44.4% (n=12). A total of 5 trials employed cluster design [21-25], whereas the remaining studies used a parallel group design. The majority of studies (51.8%) were conducted in nurses [23,24,26-37] followed by pharmacists (14.8%) [38-41], mental health therapists (11.1%) [21,25,42], dentists (7.4%) [43,44], midwives [22], physical therapists [45], patient care personnel [46], and substance abuse counselors [47]. The evaluated interventions included blended learning [35]; CD-ROM and emails [43]; computer-assisted instruction (CAI), computer-based training, or computer-mediated training [22,23,26,28,29,34,36,38,41,42,44-46]; CD-ROM [24,30-33,37,47]; PowerPoint presentation [39,40,46]; and software [21,25,27]. The duration of the intervention ranged from 50 min [28,38] to 3 months [21,25,31,46]. The intensity ranged from 15 min [44] to 2.4 h [37]. Comparison groups included no intervention [22,26,32,33,37,43-45,47], blended learning [21,25], and traditional learning [23,24,27-31,34-36,38-42,46]. Primary outcomes included knowledge in 20 studies [21-24,26-48], skills in 9 studies [22,24,31,35,37,38,41-43], attitudes in 7 studies

[26,30,35,40,41,44,45], satisfaction in 9 studies [23,25,28,33,35,36,38,40,42], and patient-centered outcomes in 2 studies [22,25].

Risk of Bias in Included Studies

We present our judgments about each risk of bias item for all included studies as (summary) percentages in Figure 2.

Figure 3 shows separate judgments about each risk of bias item for each included study.

The risk of bias was predominantly low for random sequence generation (55.5% of the studies), selective reporting, baseline comparability, and consistency in intervention delivery. The risk of bias was predominantly unclear for allocation concealment blinding of participants, personnel, or outcome assessors. A total of 12 studies (44.4%) had a high risk of attrition bias; 6 studies (22.2%) had a high risk of bias for validity and reliability of outcome measures; and 5 studies (18.5%) had a high risk of bias for baseline comparability. In total, 3 studies (11.1%) had a high risk of performance bias; and 1 study (3.7%) had a high risk of detection bias. For cRCTs, all 5 studies had a low risk of bias for selective recruitment of cluster participants.

Table 1. Characteristics of included studies

Author (year), reference, country	Population/health profession (N)	Field of study/condition/health problem	Intervention type	Control	Outcomes (measurement instrument)	Results (continuous or dichotomous)
Akar (2014) [46], Turkey	Patient care personnel (96)	Testicular cancer	PowerPoint presentation	T[a]	Knowledge (MCQ 26-items)	Mean (SD) 12.0 (1.9) vs 10.4 (3.7); *P*=.005
Albert (2006) [43], United States	Dentists (184)	Tobacco addiction	CD-ROM and email	NL[b]	1. Skills 2. Knowledge	1. *P*<.01 2. *P*<.05
Bayne (1997) [26], United States	Nurses (67)	Drug overdose	CAI[c]	NL	1. Knowledge (test 20-items) 2. Satisfaction 3. Attitude (Q[e])	1. Mean (SD) 82.1 (11.88) vs 81.1 (13.0) 2. —[d] 3. —
Beidas (2012) [42], United States	Mental health therapist (115)	Anxious children	CBL[f]	T	1. Skills (checklist) 2. Knowledge (test 20-items) 3. Satisfaction (Q)	1. Mean (SD) 17.4 (1.81) vs 17.4 (1.83) 2. Mean (SD) 3.6 (1.47) vs 4.1 (1.45) 3. Mean (SD) 50.8 (5.9) vs 53.7 (5.4); (*P*<.001)
Boh (1990) [38], United States	Pharmacists (105)	Osteoarthritis	Computer-based simulation	T	1. Knowledge (MCQ[g] 25-items) 2. Skills (simulation) 3. Satisfaction (Q)	1. Mean (SD) 76.0 (8.59) vs 65.73 (9.65); (*P*<.005) 2. Mean (SD) 32.9 (8.15) vs 26.5 (10.90) 3. —
Bredesen (2016) [27], Norway	Nurses (44)	Pressure ulcer prevention	Software	T	1. Knowledge (a. Braden scale and b. pressure ulcer classification)	a. NS[h] b. Fleiss kappa=0.20 (0.18-0.22) vs 0.27 (0.25-0.29)
Chiu (2009) [28], Taiwan	Nurses (84)	Stroke	CAI	T	1. Knowledge (Q 15-items) 2. Satisfaction (Q 16-items)	1. Mean (SD) 34.7 (2.4) vs 33.7 (5.0); (*P*=.21) 2. Mean (SD) 61.5 (8.40) vs 60.3 (7.80); (*P*=.51)
Cox (2009) [29], United States	Nurses (60)	Pressure ulcers	CBL	T	Knowledge (Single choice questionnaire)	Mean (SD) 90.3 (4.9) vs 92.9 (3.3); (*P*=.717)
de Beurs (2015) [21], Holland	Psychiatric departments (567[i])	Suicide prevention	Software (train-the-trainer)[a]	BL[j]	Knowledge (Q 15-items)	Mean (SD) 26.6 (3.1) vs 24.1 (2.3)
de Beurs 2016 [46], Holland	Psychiatric departments (881[i])	Suicide prevention	Software (train-the-trainer)[k]	BL	1. Patient-centered outcome (Beck scale 19-items) 2. Satisfaction (4-point scale)	1. Mean (SD) 4.2 (13.4) vs 4.9 (10.5) 2. Mean (SD) 6.8 (4.4) vs 6.8 (4.3)
Donyai (2015) [39], United Kingdom	Pharmacy professional (48)	Continuing professional development case scenarios	PowerPoint presentation	T	Knowledge (score)	Mean difference 9.9; 95% CI 0.4 to 19.3; (*P*=.04)
Ebadi (2015) [30], Iran	Nurses (90)	Biological incidents	CD-ROM	T	1. Knowledge (MCQ 34-items) 2. Attitude (visual analogue scale 0-100)	1. Mean (SD)=24.3 (5.1) vs 13.9 (3.2) (*P*<.001) 2. Mean (SD) 81.59 (15.21) vs 54.4 (20.24); (*P*<.001)
Gasko (2012) [31], United States	Nurse anesthetists (29)	Regional anesthesia	CD-ROM	T	Skills (Q 16 criteria)	Mean (SD) 33 (7) vs 35 (10); (*P*<.05)
Hsieh (2006) [44], United States	Dentists (174)	Domestic violence	CBL	NL	1. Knowledge (Q 24-items) 2. Attitude	1. (*P*<.01) 2. (*P*<.01)

Author (year), reference, country	Population/health profession (N)	Field of study/condition/health problem	Intervention type	Control	Outcomes (measurement instrument)	Results (continuous or dichotomous)
Ismail (2013) [22], United Kingdom	Midwives (25)	Perineal trauma	CBL	NL	Patient-centered outcomes	Delta=0.7%; 95% CI −10.1 to 11.4; (*P*=.89)
Javadi (2015) [40], Iran	Pharmacists (71)	Contraception and sexual dysfunctions	PowerPoint presentation	T	1. Knowledge (MCQ 23-items)	1. Mean (SD) 68.46 (16.60) vs 50.75 (17.58); (*P*<.001)
					2. Satisfaction (Q 5-items)	2. —
					3. Attitude (scale 14-items)	3. Median 28 vs 27; (*P*=.18)
Lawson (1991) [41], United States	Pharmacists (50)	Financial management	CBL	T	1. Skills (Q 25-items)	1. Mean (SD) 15.63 (3.37) vs 16.04 (3.35)
					2. Attitude	2. (*P*=.082)
Liu (2014) [32], Taiwan	Psychiatric nurses (216)	Case management	CD-ROM	NL	Knowledge (MCQ 20-items)	Delta= 0.37; 95% CI –3.3 to 4.0; (*P*=.84)
Liu (2014) [33], Taiwan	Nursing personnel (40)	Nursing care management	CD-ROM	NL	Knowledge (Q)	Mean (SD) 91 (8.6) vs 58 (20.4)
Moran (1991) [45], United States	Physical therapists (41)	Wound care	CAI	NL	1. Knowledge (test 13-items)	1. Mean (SD) 10.85(1.56) vs 9.05(1.77); (*P*<.004)
					2. Attitude (survey)	2. —
Padalino (2007) [34], Brazil	Nurses (49)	Quality training program	CBL	T	Knowledge (Q)	Mean (SD) 19.4(1.7) vs 17.8(3.2); (*P*=.072)
Pun (2016) [35], Hong Kong	Nurses (40)	Hemodialysis management	BL	T	1. Knowledge (MCQ and fill-in-the blank questions)	1. Mean (SD) 24 (1.03) vs 17.45 (2.74); (*P*<.001)
					2. Skills (checklist 39-items)	2. Mean (SD) 149.3 (19.42) vs 113.65 (21.23); (*P*<.001)
					3. Attitude (3-item checklist 7-point Likert scale)	3. Mean (SD) 1.83 (0.03)[l]
					4. Satisfaction (7-point Likert scale)	4. Range 2.10 to 2.75 (0.55 to 0.94)[l]
Roh (2013) [36], Korea	Nurses (38)	Advanced life support	CBL	T	Satisfaction (Q 20-items)	Mean (SD) 7.64 (1.04) vs 7.43 (1.34); (*P*=.588)
Rosen (2002) [23], United States	Nurses (173)	Mental health and aging	CBL	T	1. Knowledge (test)	1. Mean (SD) 90.0 (9.1) vs 84.0 (11.2); (*P*=.004)
					2. Satisfaction (Q)	2. (*P*<.0001)
Schermer (2011) [24], Holland	Nurses (1135)	Spirometry	CD-ROM	T	Skills (test)	OR[m] 1.2, 95% CI 0.6 to 2.5; (*P*=.663)
Schneider (2006) [37], United States	Nurses (30)	Medication administration	CD-ROM	NL	Skills (observation)	OR 0.38, 95% CI 0.19 to 0.74; (*P*=.004)

Author (year), reference, country	Population/health profession (N)	Field of study/condition/health problem	Intervention type	Control	Outcomes (measurement instrument)	Results (continuous or dichotomous)
Weingardt (2006) [47], United States	Substance abuse counselor (166)	Substance abuse	CD-ROM	NL	Knowledge (MCQ)	(*P*<.01)

[a] T: traditional.

[b] NL: no learning.

[c] CAI: computer-assisted instruction.

[d] —: not reported.

[e] Q: questionnaire.

[f] CBL: computer-based learning.

[g] MCQ: multiple choice questionnaire.

[h] NS: not significant.

[i] Total number of patients or professions.

[j] BL: blended learning.

[k] Intervention also included blended learning (Web-based plus traditional learning).

[l] Data for the intervention group only.

[m] OR: odds ratio.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

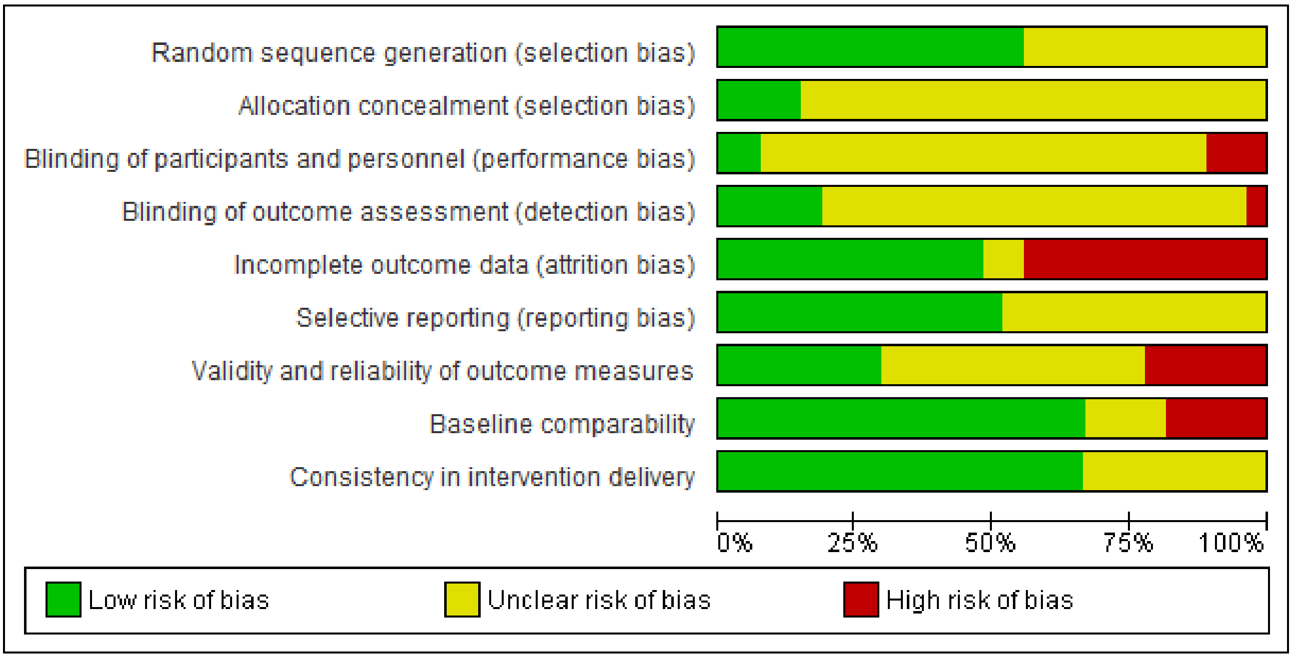
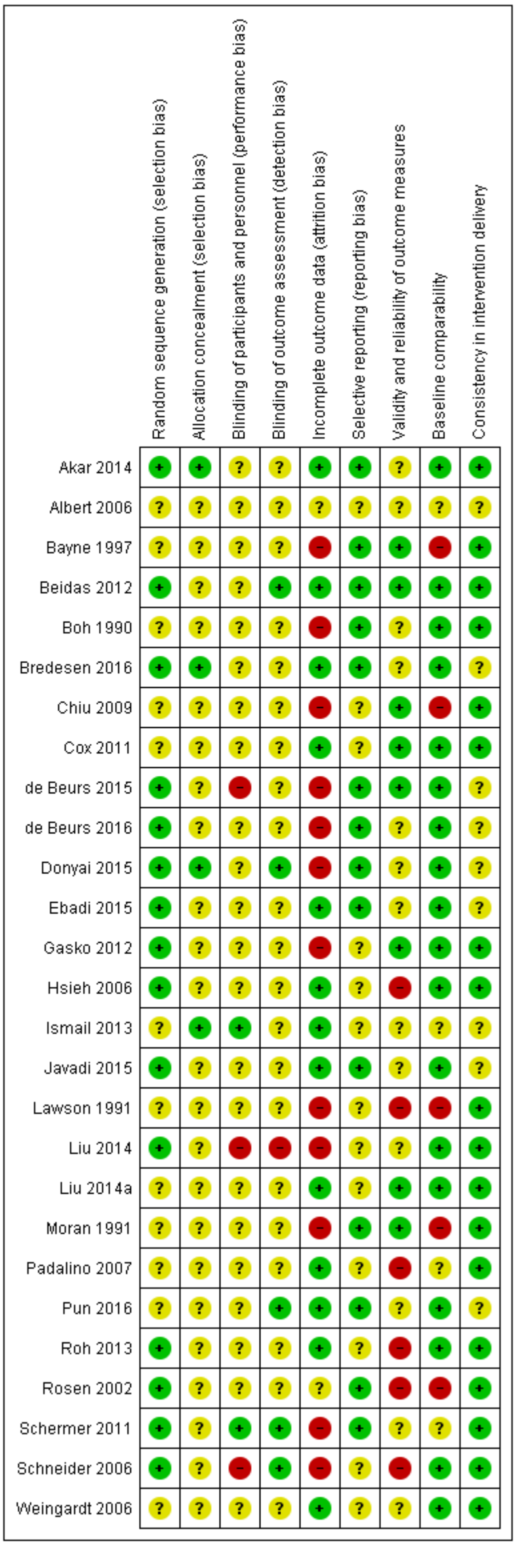


Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Validity and reliability of outcome measures	Baseline comparability	Consistency in intervention delivery
Akar 2014	+	+	?	?	+	+	?	+	+
Albert 2006	?	?	?	?	?	?	?	?	?
Bayne 1997	?	?	?	?	−	+	+	−	+
Beidas 2012	+	?	?	+	+	+	+	+	+
Boh 1990	?	?	?	?	−	+	?	+	+
Bredesen 2016	+	+	?	?	+	+	?	+	?
Chiu 2009	?	?	?	?	−	?	+	−	+
Cox 2011	?	?	?	?	+	?	+	+	+
de Beurs 2015	+	?	−	?	−	+	+	+	?
de Beurs 2016	+	?	?	?	−	+	?	+	?
Donyai 2015	+	+	?	+	−	+	?	+	?
Ebadi 2015	+	?	?	?	+	+	?	+	?
Gasko 2012	+	?	?	?	−	?	+	+	+
Hsieh 2006	+	?	?	?	+	?	−	+	+
Ismail 2013	?	+	+	?	+	?	?	?	?
Javadi 2015	+	?	?	?	+	+	?	+	?
Lawson 1991	?	?	?	?	−	?	−	−	+
Liu 2014	+	?	−	−	−	?	?	+	+
Liu 2014a	?	?	?	?	+	?	+	+	+
Moran 1991	?	?	?	?	−	+	+	−	+
Padalino 2007	?	?	?	?	+	?	−	?	+
Pun 2016	?	?	?	+	+	+	?	+	?
Roh 2013	+	?	?	?	+	?	−	+	+
Rosen 2002	+	?	?	?	?	+	−	−	+
Schermer 2011	+	?	+	+	−	+	?	?	+
Schneider 2006	+	?	−	+	−	?	−	+	+
Weingardt 2006	?	?	?	?	+	?	?	+	+

Effects of Interventions

Offline Digital Education (CD-ROM) Versus No Intervention or Traditional Learning

Primary Outcomes

Knowledge

A meta-analysis of 3 studies [30,32,33] considered to be homogeneous enough found that compared with no intervention, offline digital education (CD-ROM) may increase knowledge in nurses (SMD 1.88; 95% CI 1.14 to 2.62; low certainty evidence, Figure 4). There was a substantial level of heterogeneity of the pooled studies (Tau^2=.34; χ^2=9.90; P=.007; I^2=80%; low certainty evidence).

A total of 2 studies did not report sufficient data that could be included in the meta-analysis. Weingardt [47] reported that compared with no intervention, CD-ROM probably improves substance abuse counselors' knowledge (*P*<.01; moderate certainty evidence). Albert [43] reported that compared with no intervention, CD-ROM and email may slightly improve dentists' knowledge (*P*<.05; low certainty evidence).

Skills

Schneider [37] reported an increase in nurses' skills (decreased core 1 error rates) between baseline and postintervention periods in the intervention group (OR 0.38, 95% CI 0.19 to 0.74; *P*=.004; low certainty evidence). Albert [43] reported that compared with no intervention, the offline digital education (CD-ROM and email) intervention may slightly improve dentists' skills (*P*<.01; low certainty evidence). Gasko [31] reported that the CD-ROM intervention may have little or no effect on nurse anesthetists' skills compared with traditional learning (mean 33 [SD 7] vs mean 35 [SD 10]; low certainty evidence). Schermer [24] reported that compared with traditional training (joint baseline workshop), CD-ROM may slightly improve the rate of adequate tests (32.9% vs 29.8%; OR 1.2, 95% CI 0.6 to 2.5; *P*=.663; low certainty evidence).

Satisfaction

Liu [33] reported that 87% of participants in the CD-ROM groups agreed or strongly agreed that the program was flexible (mean 4.28; low certainty evidence). There was no comparison group for this outcome. For a summary of the effects of these comparisons on knowledge, skills, and satisfaction, see SoF in Multimedia Appendix 4.

Offline Digital Education (Computer-Assisted Training) Versus No Intervention or Traditional Learning

Primary Outcomes

Knowledge

A meta-analysis of 2 studies [26,45] considered to be homogeneous enough found that compared with no intervention, the effects of offline digital education (computer-assisted training [CAT]) on nurses and physical therapists' knowledge were uncertain (SMD 0.55; 95% CI –0.39 to 1.50; very low certainty evidence; Figure 5).

A substantial level of heterogeneity of the pooled studies was detected (Tau^2=.33; χ^2=3.40; *P*=.07; I^2=71%). One study [44] did not present data that could be included in the meta-analysis for this outcome. Hsieh reported that compared with no intervention, offline digital education may improve dentists' knowledge (*P*<.01; low certainty evidence).

Beidas [42] reported that compared with routine training, offline digital education (computer training) may have little or no effect on community mental health therapists' knowledge postintervention (mean 17.45 [SD 1.83] vs mean 17.42 [SD 1.81]; *P*=.26; low certainty evidence). Boh [38] found that compared with traditional learning, an intervention (audio cassette and microcomputer simulation) may improve pharmacists' knowledge postintervention (mean 65.7 [SD 9.6] vs mean 76 [SD 8.5]; *P*<.005; low certainty evidence). Chiu [28] reported that compared with traditional training, offline digital education (CAI) may slightly improve nurses' knowledge at 4 weeks (mean 33.7 [SD 5.0] vs mean 34.7 [SD 2.4]; *P*=.21; low certainty evidence). Cox [29] found that compared with traditional training, offline digital education (computer-based learning) may have little or no effect on nurses' knowledge postintervention (mean 92.9 [SD 3.3] vs mean 90.3 [SD 4.9]; *P*=.717; low certainty evidence). Padalino [34] reported that compared with traditional classroom training, offline digital education (computer-mediated training) may slightly improve nurses' knowledge postintervention (mean 17.8 [SD 3.2] vs mean 19.4 [SD 1.7]; *P*=.072; low certainty evidence). Rosen [23] found that compared with usual education, offline digital education (computer-based training) may improve nurses' knowledge at 6 months (mean 84 [SD 11.2] vs mean 90 [SD 9.1]; *P*=.004; low certainty evidence). Taken together, these results suggest that computer-assisted interventions may slightly improve various health professionals' knowledge, but the quality of evidence was low and results were mixed.

Skills

A meta-analysis of 4 studies [35,38,41,42] found that compared with traditional training, the effects of CAT on skills in community (mental health) therapists, nurses, and pharmacists were uncertain (SMD 0.45; 95% CI –0.35 to 1.25; very low certainty evidence; Figure 6). Heterogeneity of the pooled studies was considerable (Tau^2=.58; χ^2=25.06; *P*<.0001; I^2=88%).

Attitudes

In Moran [45], 93% of respondents reported a strong agreement or an agreement with the statement that computer-assisted instructions were helpful. There was no comparison group for this outcome (low certainty evidence). Hsieh [44] reported that compared with no intervention, the computer-based tutorial group may improve dentists' attitudes (*P*<.01; low certainty evidence). Lawson [41] found that compared with traditional education, offline digital education may have little or no effect on participants' attitudes concerning expected helpfulness (*P*=.082; low certainty evidence).

Figure 4. Forest plot of comparison: Offline digital education (CD-ROM) versus no intervention, outcome: Knowledge.

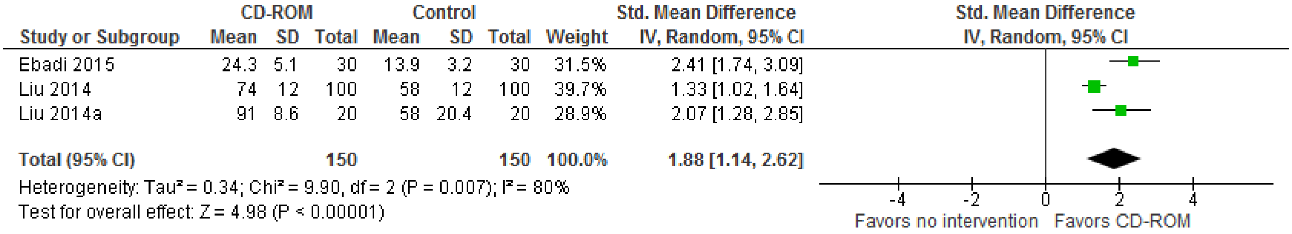
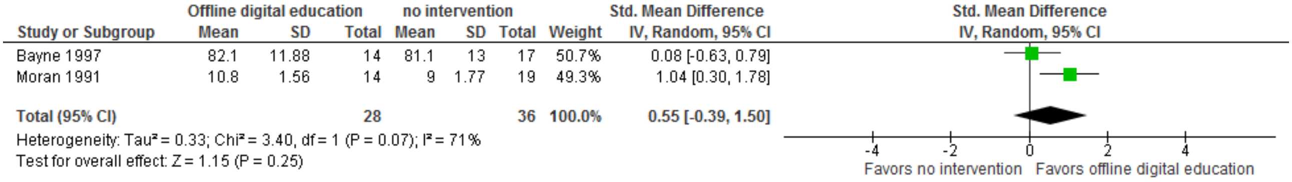


Figure 5. Forest plot of comparison: Offline digital education (computer-assisted training) versus no intervention, outcome: Knowledge.

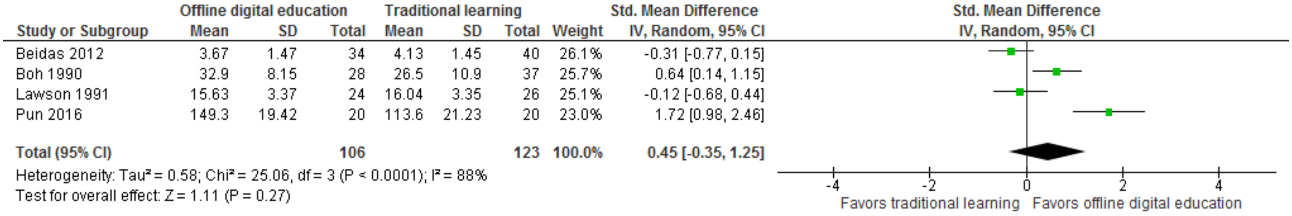


Study or Subgroup	Offline digital education Mean	SD	Total	no intervention Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bayne 1997	82.1	11.88	14	81.1	13	17	50.7%	0.08 [-0.63, 0.79]
Moran 1991	10.8	1.56	14	9	1.77	19	49.3%	1.04 [0.30, 1.78]
Total (95% CI)			**28**			**36**	**100.0%**	**0.55 [-0.39, 1.50]**

Heterogeneity: Tau² = 0.33; Chi² = 3.40, df = 1 (P = 0.07); I² = 71%

Test for overall effect: Z = 1.15 (P = 0.25)

Figure 6. Forest plot of comparison: Offline digital education (computer-assisted training) versus traditional learning, outcome: Skills.



Study or Subgroup	Offline digital education Mean	SD	Total	Traditional learning Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Beidas 2012	3.67	1.47	34	4.13	1.45	40	26.1%	-0.31 [-0.77, 0.15]
Boh 1990	32.9	8.15	28	26.5	10.9	37	25.7%	0.64 [0.14, 1.15]
Lawson 1991	15.63	3.37	24	16.04	3.35	26	25.1%	-0.12 [-0.68, 0.44]
Pun 2016	149.3	19.42	20	113.6	21.23	20	23.0%	1.72 [0.98, 2.46]
Total (95% CI)			**106**			**123**	**100.0%**	**0.45 [-0.35, 1.25]**

Heterogeneity: Tau² = 0.58; Chi² = 25.06, df = 3 (P < 0.0001); I² = 88%

Test for overall effect: Z = 1.11 (P = 0.27)

Satisfaction

A meta-analysis of 4 studies [25,28,36,42] considered to be homogeneous enough found that compared with traditional training, offline digital education may have little effect or no difference on satisfaction scores in nurses and mental health therapists (SMD –0.07; 95% CI –0.42 to 0.28; low certainty evidence; Figure 7). A moderate level of heterogeneity of the pooled studies was detected (Tau^2=.05; χ^2=.10; *P*=.16; I^2=41%).

A total of 2 studies [23,38] were not included in the meta-analysis for this outcome as they did not report a sufficient amount of data for pooling. Boh [38] found that compared with traditional learning, offline digital education (audio cassette and microcomputer simulation) may have little or no effect on pharmacists' satisfaction postintervention (low certainty evidence). Rosen [23] found that compared with usual education, offline digital education (computer-based training) may improve nurses' satisfaction at 6 months (*P*<.0001; low certainty evidence).

Patient-Centered Outcomes

Ismail [22] reported that compared with no intervention, offline digital education may have little or no effect on the average percentage of women reporting perineal pain on sitting and walking at 10 to 12 days (mean difference [MD]=0.7%; 95% CI −10.1 to 11.4; *P*=.89; low certainty evidence).

Secondary Outcomes

Only 1 study [27] mentioned the costs of offline digital education. Bayne and Bindler [27] reported the costs as US $54 per participant in the computer-assisted group compared with US $23 per participant in the no intervention control group. For a summary of the effects of these comparisons on all outcomes, see SoF Multimedia Appendix 4.

Offline Digital Education (Software, PowerPoint) Versus Blended Learning or Traditional Learning

Primary Outcomes

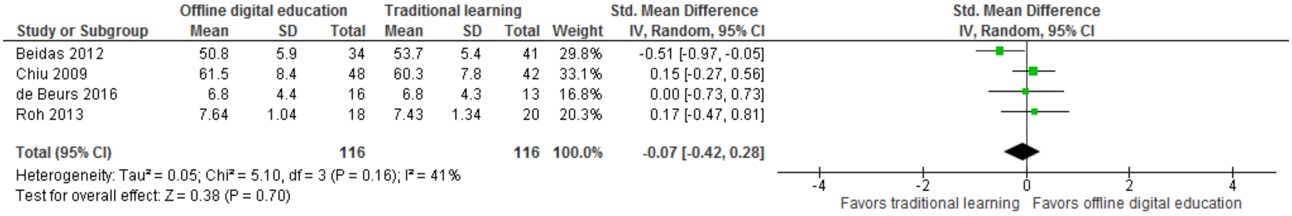
Knowledge

A meta-analysis of 2 studies [40,46] considered to be homogeneous enough found that compared with traditional learning, a PowerPoint presentation may improve the knowledge of patient care personnel and pharmacists (SMD 0.76; 95% CI 0.29 to 1.23; low certainty evidence; Figure 8). A considerable level of heterogeneity of the pooled studies was detected (Tau^2=.06; χ^2=2.19; *P*=.14; I^2=54%).

One study did not report sufficient data to be included in the meta-analysis. Donyai [39] reported that compared with traditional learning, a PowerPoint presentation may improve pharmacy professionals' knowledge (MD=9.9; 95% CI 0.4 to 19.3; *P*=.04).

de Beurs [21] reported that compared with blended learning, offline digital education (software) may improve mental health professionals' knowledge (mean [SD] 26.6 (3.1) vs 24.1 (2.3); *P*<.001; low certainty evidence).

Figure 7. Forest plot of comparison: Offline digital education (computer-assisted training) versus traditional learning, outcome: Satisfaction.

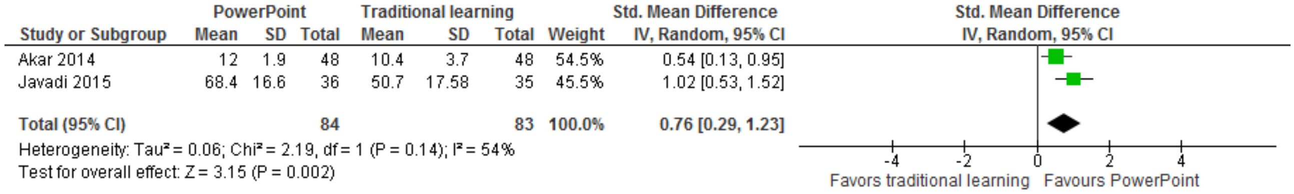


Study or Subgroup	Offline digital education Mean	SD	Total	Traditional learning Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Beidas 2012	50.8	5.9	34	53.7	5.4	41	29.8%	-0.51 [-0.97, -0.05]
Chiu 2009	61.5	8.4	48	60.3	7.8	42	33.1%	0.15 [-0.27, 0.56]
de Beurs 2016	6.8	4.4	16	6.8	4.3	13	16.8%	0.00 [-0.73, 0.73]
Roh 2013	7.64	1.04	18	7.43	1.34	20	20.3%	0.17 [-0.47, 0.81]
Total (95% CI)			**116**			**116**	**100.0%**	**-0.07 [-0.42, 0.28]**

Heterogeneity: Tau² = 0.05; Chi² = 5.10, df = 3 (P = 0.16); I² = 41%

Test for overall effect: Z = 0.38 (P = 0.70)

Figure 8. Forest plot of comparison: Offline digital education (PowerPoint) versus traditional learning, outcome: Knowledge.



Study or Subgroup	PowerPoint Mean	SD	Total	Traditional learning Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Akar 2014	12	1.9	48	10.4	3.7	48	54.5%	0.54 [0.13, 0.95]
Javadi 2015	68.4	16.6	36	50.7	17.58	35	45.5%	1.02 [0.53, 1.52]
Total (95% CI)			**84**			**83**	**100.0%**	**0.76 [0.29, 1.23]**

Heterogeneity: Tau² = 0.06; Chi² = 2.19, df = 1 (P = 0.14); I² = 54%

Test for overall effect: Z = 3.15 (P = 0.002)

Satisfaction

de Beurs [25] reported that compared with blended learning, offline digital education (software) may have little effect or no difference on patients' satisfaction at 3 months (mean [SD] 6.8 (4.4) vs 6.8 (4.3); low certainty evidence).

Patient-Centered Outcomes

de Beurs [25] reported that compared with blended learning, offline digital education (software) may have little effect or no difference on patients' suicidal ideation at 3 months (mean [SD] 4.2 (13.4) vs 4.9 (10.5); low certainty evidence). For a summary of the effects of these comparisons on all outcomes, see SoF Multimedia Appendix 4.

There was not enough data included in any of the pooled analyses to allow sensitivity analyses to be conducted. Similarly, given the small number of trials contributing data to outcomes within different comparisons in this review, a formal assessment of potential publication bias was not feasible.

Discussion

We summarized and critically evaluated evidence for effectiveness of offline digital education for improving knowledge, skills, attitudes, satisfaction, and patient-centered outcomes in postgraduate health professions except medical doctors. A total of 27 studies with 4618 participants met the eligibility criteria. We found highly diverse studies in different professions and evidence to support the effectiveness of certain types of offline digital education such as CD-ROM and PowerPoint compared with no intervention or traditional learning in improving learners' knowledge. For other outcomes (and comparators), the evidence was less compelling in improving learners' skills, attitudes, satisfaction, and patient-related outcomes.

Overall Completeness and Applicability of Evidence

We identified 4 studies from upper middle-income countries (Brazil, Iran, and Turkey), and the remaining studies were conducted in high-income countries (Hong Kong, Korea, the Netherlands, Norway, the United Kingdom, and the United States). Only 4 studies (15%) were conducted during the 1990s and the remaining studies were from 2000 onward. In 15 studies (55.5%), information about the frequency of the interventions was missing, thereby often making it difficult to analyze in depth and interpret the findings. Similarly, economic evaluations of the interventions were missing in 26 (96%) studies.

Quality of the Evidence

Overall, the quality of evidence was low or very low. We assessed the quality of evidence using the Grading of Recommendations, Assessment, Development, and Evaluations system and presented the findings in SoF Multimedia Appendix 4 for all comparisons. The reasons for downgrading the evidence most commonly pertained to the high risk of bias. For instance, only 13 (48.1%) of the studies reported complete outcome data. Reducing the dropout rate might reduce the risk of attrition bias and further improve the quality of the studies. Only 8 studies (29.6%) had a low risk of bias for validity and reliability of outcome measures. This issue of nonvalidated measurement tools has repeatedly been raised and is paramount to advance the field [49]. Only 2 studies (7.4%) adequately described blinding of participants and personnel. As with many educational interventions, blinding of participants or personnel might prove challenging. However, we highlighted the need for more adequate descriptions of masking to further reduce the risk of performance bias and allow clearer judgments to be made. We also downgraded the overall quality of evidence for inconsistency (where there was a high level of heterogeneity, ie, I^2>50%). Overall, there was a moderate-to-considerable level of heterogeneity of meta-analyses (I^2 range 41% to 88%); and 4 (out of 5) meta-analyses had I^2>50%. More reasons for downgrading included indirectness (we downgraded once for 1 outcome only—where there were differences in the population used). Participants were not homogeneous and ranged from nurses, pharmacists, mental health therapists, dentists, midwives and obstetricians, physical therapists, patient care personnel to substance abuse counselors. Other sources of indirectness also stemmed from heterogeneous interventions (their duration, frequency, and intensity), comparison groups, and outcome assessment tools ranging from multiple choice or single choice questionnaires, tests, observations, checklists, scales, surveys, visual analogue scales, and simulations. Finally, we also downgraded for imprecision where the sample size was small. The included studies also failed to provide details of sample size and power calculations and may have therefore been underpowered and unable to detect change in learning outcomes.

Strengths and Limitations of the Review

This systematic review has several important strengths that include comprehensive searches without any language limitations, robust screening, data extraction and risk of bias assessments, and a critical appraisal of the evidence. However, some limitations must be acknowledged while interpreting the results of this study. First, we considered subgroup analyses to be unfeasible because of the insufficient number of studies under the respective outcomes and professional groups. However, we minimized potential biases in the review process and maintained its internal validity by strictly adhering to the guidelines outlined by Higgins et al [18].

Agreements and Disagreements With Other Studies or Reviews

A review by Al-Jewair [50] found some evidence to support the effectiveness of computer-assisted learning in improving knowledge gains in undergraduate or postgraduate orthodontic students' or orthodontic educators' knowledge, but no definite conclusions were reached; and future research was recommended. Rosenberg [51] concluded that computer-aided learning is as effective as other methods of teaching and can be used as an adjunct to traditional education or as a means of self-instruction of dental students. Based on 4 mixed-results RCTs, Rosenberg [52] was unable to reach any conclusions on knowledge gains and recommended more high-quality trials evaluating the effectiveness of computer-aided learning in orthodontics. However, we are familiar with newer technologies being currently evaluated for the same outcomes; MOOCs or mLearning can play a very important role in health professions education such as improving clinical knowledge and promoting lifelong learning [53-54]. We are also aware of recent reviews, which reached similar conclusions [55-62]. For example, digital education seems to be at least as effective (and sometimes more effective) as traditional education in improving dermatology, diabetes management, or smoking cessation–related skills and knowledge [58,61,62]. Most of these reviews, however, stressed the inconclusiveness of overall findings mainly because of the low certainty of the evidence.

Conclusions

Offline digital education may potentially play a role in the education of health professionals, especially in LMICs, where there is a lack of access to Web-based digital education for a variety of reasons, including cost; and there is some evidence to support the effectiveness of these interventions in improving the knowledge of health professionals. However, because of the existing gaps in the evidence base, including limited evidence for other outcomes; lack of subgroup analyses, for example, CD-ROM or PowerPoint; low and very low quality of the evidence, the overall findings are inconclusive. More research especially evaluating patient-centered outcomes, costs, and safety (adverse effects); involving those subgroups; and originating from LMICs is needed. Such research should be adequately powered, be underpinned by learning theories, use valid and reliable outcome measures, and blind outcome assessors.

Acknowledgments

This review was conducted in collaboration with the World Health Organization, Health Workforce Department. We would like to thank Carl Gornitzki, GunBrit Knutssön, and Klas Moberg from the University Library, Karolinska Institutet, Sweden, for developing the search strategy. We gratefully acknowledge funding from Nanyang Technological University Singapore, Singapore, eLearning for health professionals education grant.

Authors' Contributions

JC and PP conceived the idea for the review. PP and MMB wrote the review. BMK, MS, UD, MK, and AS extracted data and appraised for risk of bias, and all authors commented on the review and made revisions following the first draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[DOCX File, 17KB - jmir_v21i4e12968_app1.docx]

Multimedia Appendix 2

Characteristics of excluded studies.

[DOCX File, 13KB - jmir_v21i4e12968_app2.docx]

Multimedia Appendix 3

Results of the included studies.

[DOCX File, 47KB - jmir_v21i4e12968_app3.docx]

Multimedia Appendix 4

Summary of findings tables.

[DOCX File, 20KB - jmir_v21i4e12968_app4.docx]

References

1. Anand S, Bärnighausen T. Human resources and health outcomes: cross-country econometric study. Lancet 2004;364(9445):1603-1609. [doi: 10.1016/S0140-6736(04)17313-3] [Medline: 15519630]
2. Campbell J, Dussault G, Buchan J, Pozo-Martin F, Guerra Arias M, Leone C, et al. World Health Organization. 2013. A universal truth: no health without a workforce URL: https://www.who.int/workforcealliance/knowledge/resources/GHWA-a_universal_truth_report.pdf?ua=1 [WebCite Cache ID 77g8T1kVD]
3. Chen L. Striking the right balance: health workforce retention in remote and rural areas. Bull World Health Organ 2010 May;88(5):323, A [FREE Full text] [doi: 10.2471/BLT.10.078477] [Medline: 20461215]
4. Moszynski P. One billion people are affected by global shortage of healthcare workers. Br Med J 2011 Feb 1;342:d696. [doi: 10.1136/bmj.d696] [Medline: 21285190]
5. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
6. Nkomazana O, Mash R, Shaibu S, Phaladze N. Stakeholders' perceptions on shortage of healthcare workers in primary healthcare in Botswana: focus group discussions. PLoS One 2015;10(8):e0135846 [FREE Full text] [doi: 10.1371/journal.pone.0135846] [Medline: 26284617]
7. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education for health professions: methods for overarching evidence syntheses. J Med Internet Res 2019 Dec 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
8. DeBate RD, Severson HH, Cragun DL, Gau JM, Merrell LK, Bleck JR, et al. Evaluation of a theory-driven e-learning intervention for future oral healthcare providers on secondary prevention of disordered eating behaviors. Health Educ Res 2013 Jun;28(3):472-487 [FREE Full text] [doi: 10.1093/her/cyt050] [Medline: 23564725]
9. Kahiigi EK, Ekenberg L, Hansson H, Tusubira FF, Danielson M. Exploring the e-Learning state of art. EJEL 2008;6(2):77-87 [FREE Full text]
10. Rasmussen K, Belisario JM, Wark PA, Molina JA, Loong SL, Cotic Z, et al. Offline eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010405 [FREE Full text] [doi: 10.7189/jogh.04.010405] [Medline: 24976964]
11. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
12. Greenhalgh T. Computer assisted learning in undergraduate medical education. Br Med J 2001 Jan 6;322(7277):40-44 [FREE Full text] [Medline: 11141156]
13. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
14. Lam-Antoniades M, Ratnapalan S, Tait G. Electronic continuing education in the health professions: an update on evidence from RCTs. J Contin Educ Health Prof 2009;29(1):44-51. [doi: 10.1002/chp.20005] [Medline: 19288566]
15. Adams AM. Pedagogical underpinnings of computer-based learning. J Adv Nurs 2004 Apr;46(1):5-12. [doi: 10.1111/j.1365-2648.2003.02960.x] [Medline: 15030437]
16. Brandt MG, Davies ET. Visual-spatial ability, learning modality and surgical knot tying. Can J Surg 2006 Dec;49(6):412-416 [FREE Full text] [Medline: 17234070]
17. Mahan JD, Stein DS. Teaching adults-best practices that leverage the emerging understanding of the neurobiology of learning. Curr Probl Pediatr Adolesc Health Care 2014 Jul;44(6):141-149. [doi: 10.1016/j.cppeds.2014.01.003] [Medline: 24981663]
18. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. 2018. URL: http://www.cochrane-handbook.org/ [WebCite Cache ID 74GJXiN6p]
19. Hervatis V, Kyaw B, Semwal M, Dunleavy G, Tudor Car L, Zary N, et al. Cochrane Database System Review. 2016 Aug 10. Offline and computer-based eLearning interventions for medical students' education URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=45679 [accessed 2019-04-17] [WebCite Cache ID 77g7DUEpo]
20. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Cochrane Handbook for Systematic Reviews of Interventions. UK; 2011. Chapter 11: Presenting results and 'Summary of findings' tables URL: https://handbook-5-1.cochrane.org/chapter_11/11_presenting_results_and_summary_of_findings_tables.htm [accessed 2019-04-17] [WebCite Cache ID 77hLEEMLL]
21. Cox J, Roche S, van Wynen E. The effects of various instructional methods on retention of knowledge about pressure ulcers among critical care and medical-surgical nurses. J Contin Educ Nurs 2011 Feb;42(2):71-78. [doi: 10.3928/00220124-20100802-03] [Medline: 20704094]
22. Padalino Y, Peres HH. E-learning: a comparative study for knowledge apprehension among nurses. Rev Lat Am Enfermagem 2007;15(3):397-403 [FREE Full text] [Medline: 17653421]

23. Albert D, Gordon J, Andrews J, Severson H, Ward A, Sadowsky D. Use of electronic continuing education in dental offices for tobacco cessation (PA2-7). 2006 Presented at: Society for Research on Nicotine and Tobacco 12th Annual Meeting; February 15-18, 2006; Orlando, Florida.
24. Hsieh NK, Herzig K, Gansky SA, Danley D, Gerbert B. Changing dentists' knowledge, attitudes and behavior regarding domestic violence through an interactive multimedia tutorial. J Am Dent Assoc 2006 May;137(5):596-603. [doi: 10.14219/jada.archive.2006.0254] [Medline: 16739538]
25. Weingardt KR, Villafranca SW, Levin C. Technology-based training in cognitive behavioral therapy for substance abuse counselors. Subst Abus 2006 Sep;27(3):19-25. [doi: 10.1300/J465v27n03_04] [Medline: 17135177]
26. Ismail KM, Kettle C, Macdonald SE, Tohill S, Thomas PW, Bick D. Perineal Assessment and Repair Longitudinal Study (PEARLS): a matched-pair cluster randomized trial. BMC Med 2013 Sep 23;11:209 [FREE Full text] [doi: 10.1186/1741-7015-11-209] [Medline: 24059602]
27. Bayne T, Bindler R. Effectiveness of medication calculation enhancement methods with nurses. J Nurs Staff Dev 1997;13(6):293-301. [Medline: 9429370]
28. Bredesen IM, Bjøro K, Gunningberg L, Hofoss D. Effect of e-learning program on risk assessment and pressure ulcer classification - a randomized study. Nurse Educ Today 2016 May;40:191-197. [doi: 10.1016/j.nedt.2016.03.008] [Medline: 27125172]
29. Chiu S, Cheng K, Sun T, Chang K, Tan T, Lin T, et al. The effectiveness of interactive computer assisted instruction compared to videotaped instruction for teaching nurses to assess neurological function of stroke patients: a randomized controlled trial. Int J Nurs Stud 2009 Dec;46(12):1548-1556. [doi: 10.1016/j.ijnurstu.2009.05.008] [Medline: 19552905]
30. Gasko J, Johnson A, Sherner J, Craig J, Gegel B, Burgert J, et al. Effects of using simulation versus CD-ROM in the performance of ultrasound-guided regional anesthesia. AANA J 2012 Aug;80(4 Suppl):S56-S59. [Medline: 23248832]
31. Liu W, Chu K, Chen S. The development and preliminary effectiveness of a nursing case management e-learning program. Comput Inform Nurs 2014 Jul;32(7):343-352. [doi: 10.1097/CIN.0000000000000050] [Medline: 24633234]
32. Schneider PJ, Pedersen CA, Montanya KR, Curran CR, Harpe SE, Bohenek W, et al. Improving the safety of medication administration using an interactive CD-ROM program. Am J Health Syst Pharm 2006 Jan 1;63(1):59-64. [doi: 10.2146/ajhp040609] [Medline: 16373466]
33. Boh LE, Pitterle ME, Demuth JE. Evaluation of an Integrated Audio Cassette and Computer Based Patient Simulation Learning Program. Am J Pharm Educ 1990;54:15-22 [FREE Full text]
34. Javadi M, Kargar A, Gholami K, Hadjibabaie M, Rashidian A, Torkamandi H, et al. Didactic lecture versus interactive workshop for continuing pharmacy education on reproductive health: a randomized controlled trial. Eval Health Prof 2015 Sep;38(3):404-418. [doi: 10.1177/0163278713513949] [Medline: 24326480]
35. Lawson K, Shepherd M, Gardner V. Using a computer-simulation program and a traditional approach to teach pharmacy financial management. Am J Pharm Educ 1991;55:226-235 [FREE Full text]
36. Beidas RS, Edmunds JM, Marcus SC, Kendall PC. Training and consultation to promote implementation of an empirically supported treatment: a randomized trial. Psychiatr Serv 2012 Jul;63(7):660-665 [FREE Full text] [doi: 10.1176/appi.ps.201100401] [Medline: 22549401]
37. Moran ML. Hypertext computer-assisted instruction for geriatric physical therapists. Phys Occup Ther Geriatr 2009 Jul 28;10(2):31-53. [doi: 10.1080/J148v10n02_03]
38. Schermer TR, Akkermans RP, Crockett AJ, van Montfort M, Grootens-Stekelenburg J, Stout JW, et al. Effect of e-learning and repeated performance feedback on spirometry test quality in family practice: a cluster trial. Ann Fam Med 2011;9(4):330-336 [FREE Full text] [doi: 10.1370/afm.1258] [Medline: 21747104]
39. Ebadi A, Yousefi S, Khaghanizade M, Saeid Y. Assessment competency of nurses in biological incidents. Trauma Mon 2015 Nov;20(4):e25607 [FREE Full text] [doi: 10.5812/traumamon.25607] [Medline: 26839862]
40. Pun S, Chiang VC, Choi K. A computer-based method for teaching catheter-access hemodialysis management. Comput Inform Nurs 2016 Oct;34(10):476-483. [doi: 10.1097/CIN.0000000000000262] [Medline: 27315367]
41. Roh YS, Lee WS, Chung HS, Park YM. The effects of simulation-based resuscitation training on nurses' self-efficacy and satisfaction. Nurse Educ Today 2013 Feb;33(2):123-128. [doi: 10.1016/j.nedt.2011.11.008] [Medline: 22153054]
42. Rosen J, Mulsant BH, Kollar M, Kastango KB, Mazumdar S, Fox D. Mental health training for nursing home staff using computer-based interactive video: a 6-month randomized trial. J Am Med Dir Assoc 2002;3(5):291-296. [doi: 10.1097/01.JAM.0000027201.90817.21] [Medline: 12807615]
43. de Beurs DP, de Groot MH, de Keijser J, van Duijn E, de Winter RF, Kerkhof AJ. Evaluation of benefit to patients of training mental health professionals in suicide guidelines: cluster randomised trial. Br J Psychiatry 2016 Dec;208(5):477-483. [doi: 10.1192/bjp.bp.114.156208] [Medline: 26678866]
44. Liu W, Rong J, Liu C. Using evidence-integrated e-learning to enhance case management continuing education for psychiatric nurses: a randomised controlled trial with follow-up. Nurse Educ Today 2014 Nov;34(11):1361-1367. [doi: 10.1016/j.nedt.2014.03.004] [Medline: 24713125]
45. Donyai P, Alexander AM. Training on the use of a bespoke continuing professional development framework improves the quality of CPD records. Int J Clin Pharm 2015 Dec;37(6):1250-1257. [doi: 10.1007/s11096-015-0202-4] [Medline: 26415738]

46. de Beurs DP, de Groot MH, de Keijser J, Mokkenstorm J, van Duijn E, de Winter RF, et al. The effect of an e-learning supported Train-the-Trainer programme on implementation of suicide guidelines in mental health care. J Affect Disord 2015 Apr 1;175:446-453 [FREE Full text] [doi: 10.1016/j.jad.2015.01.046] [Medline: 25679199]
47. Akar SZ, Bebiş H. Evaluation of the effectiveness of testicular cancer and testicular self-examination training for patient care personnel: intervention study. Health Educ Res 2014 Dec;29(6):966-976. [doi: 10.1093/her/cyu055] [Medline: 25248831]
48. Monteiro RS, Letchworth P, Evans J, Anderson M, Duffy S. Obstetric emergency training in a rural South African hospital: simple measures improve knowledge. 2016 May 11 Presented at: Obstetric Anaesthesia 2016; May 19-20, 2016; Manchester Central Convention Complex, UK.
49. Law GC, Apfelbacher C, Posadzki PP, Kemp S, Tudor Car L. Choice of outcomes and measurement instruments in randomised trials on eLearning in medical education: a systematic mapping review protocol. Syst Rev 2018 Dec 17;7(1):75 [FREE Full text] [doi: 10.1186/s13643-018-0739-0] [Medline: 29776434]
50. Al-Jewair TS, Azarpazhooh A, Suri S, Shah PS. Computer-assisted learning in orthodontic education: a systematic review and meta-analysis. J Dent Educ 2009 Jun;73(6):730-739 [FREE Full text] [Medline: 19491350]
51. Rosenberg H, Grad HA, Matear DW. The effectiveness of computer-aided, self-instructional programs in dental education: a systematic review of the literature. J Dent Educ 2003 May;67(5):524-532 [FREE Full text] [Medline: 12809187]
52. Rosenberg H, Sander M, Posluns J. The effectiveness of computer-aided learning in teaching orthodontics: a review of the literature. Am J Orthod Dentofacial Orthop 2005 May;127(5):599-605. [doi: 10.1016/j.ajodo.2004.02.020] [Medline: 15877041]
53. Zhou Y, Yang Y, Liu L, Zeng Z. [Effectiveness of mobile learning in medical education: a systematic review]. Nan Fang Yi Ke Da Xue Xue Bao 2018 Nov 30;38(11):1395-1400 [FREE Full text] [doi: 10.12122/j.issn.1673-4254.2018.11.20] [Medline: 30514692]
54. Liyanagunawardena TR, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014 Aug 14;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
55. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the Digital Health Education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
56. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Car LT. Mobile digital education for health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
57. Car LT, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945 [FREE Full text] [doi: 10.2196/12945] [Medline: 30816846]
58. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Car LT. Digital education for health professions on smoking cessation management: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 4;21(3):e13000 [FREE Full text] [doi: 10.2196/13000] [Medline: 30829576]
59. Wahabi HA, Esmaeil SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical doctors' offline computer-assisted digital education: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 1;21(3):e12998 [FREE Full text] [doi: 10.2196/12998] [Medline: 30821689]
60. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Online digital education for postregistration training of medical doctors: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 25;21(2):e13269 [FREE Full text] [doi: 10.2196/13269] [Medline: 30801252]
61. Huang Z, Semwal M, Lee SY, Tee M, Ong W, Tan WS, et al. Digital health professions education on diabetes management: systematic review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 21;21(2):e12997 [FREE Full text] [doi: 10.2196/12997] [Medline: 30789348]
62. Xu X, Posadzki PP, Lee GE, Car J, Smith HE. Digital education for health professions in the field of dermatology: a systematic review by Digital Health Education collaboration. Acta Derm Venereol 2019 Feb 1;99(2):133-138 [FREE Full text] [doi: 10.2340/00015555-3068] [Medline: 30320871]

Abbreviations

CAI: computer-assisted instruction
CAT: computer-assisted training
cRCT: cluster randomized controlled trial
LMIC: low- and middle-income country
MD: mean difference
mLearning: mobile learning
MOOC: massive open online course
OR: odds ratio
RCT: randomized controlled trial

SMD: standardized mean difference
SoF: Summary of Findings

Edited by A Marusic; submitted 29.11.18; peer-reviewed by KC Chu, D Adeloye; comments to author 12.12.18; revised version received 21.01.19; accepted 02.04.19; published 24.04.19.

Please cite as:
Posadzki P, Bala MM, Kyaw BM, Semwal M, Divakar U, Koperny M, Sliwka A, Car J
Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2019;21(4):e12968
URL: http://www.jmir.org/2019/4/e12968/
doi:10.2196/12968
PMID:31017584

Original Paper

Clinicians' Concerns About Mobile Ecological Momentary Assessment Tools Designed for Emerging Psychiatric Problems: Prospective Acceptability Assessment of the MEmind App

Christophe Lemey[1,2,3], MD; Mark Erik Larsen[4], PhD; Jordan Devylder[5], PhD; Philippe Courtet[6,7,8], MD, PhD; Romain Billot[3], PhD; Philippe Lenca[3], PhD; Michel Walter[1,2], MD, PhD; Enrique Baca-García[9,10,11,12,13,14,15], MD, PhD; Sofian Berrouiguet[1,3], MD

[1]EA 7479 SPURBO, Université de Bretagne Occidentale, Brest, France
[2]URCI Mental Health Department, Brest Medical University Hospital, Brest, France
[3]IMT Atlantique, Lab-STICC, F-29238 Brest, Brest, France
[4]Black Dog Institute, University of New South Wales, Sydney, Australia
[5]Graduate School of Social Service, Fordham University, New York, NY, United States
[6]Inserm U1061, La colombière Hospital, University of Montpellier, Montpellier, France
[7]Department of Emergency Psychiatry and Acute Care, CHU Montpellier, University of Montpellier, Montpellier, France
[8]Fondamental Foundation, Créteil, France
[9]Carlos III Institute Of Health, CIBERSAM (Centro de Investigation en Salud Mental), Madrid, Spain
[10]Department of Psychiatry, Universitad Catolica Del Maule, Talca, Chile
[11]Department of Psychiatry, General Hospital of Villalba, Madrid, Spain
[12]Department of Psychiatry, University Hospital Rey Juan Carlos, Mostoles, Spain
[13]Deparment of Psychiatry, University Hospital Infanta Elena, Valdemoro, Spain
[14]Psychiatry Department, Autonoma University, Madrid, Spain
[15]Department of Psychiatry, IIS-Jimenez Diaz Fondation, Madrid, Spain

Corresponding Author:
Christophe Lemey, MD
URCI Mental Health Department
Brest Medical University Hospital
Route de Ploudalmézeau
Brest,
France
Phone: 33 619211032
Email: christophe.lemey@chu-brest.fr

Abstract

Background: Many mental disorders are preceded by a prodromal phase consisting of various attenuated and unspecific symptoms and functional impairment. Electronic health records are generally used to capture these symptoms during medical consultation. Internet and mobile technologies provide the opportunity to monitor symptoms emerging in patients' environments using ecological momentary assessment techniques to support preventive therapeutic decision making.

Objective: The objective of this study was to assess the acceptability of a Web-based app designed to collect medical data during appointments and provide ecological momentary assessment features.

Methods: We recruited clinicians at 4 community psychiatry departments in France to participate. They used the app to assess patients and to collect data after viewing a video of a young patient's emerging psychiatric consultation. We then asked them to answer a short anonymous self-administered questionnaire that evaluated their experience, the acceptability of the app, and their habit of using new technologies.

Results: Of 24 practitioners invited, 21 (88%) agreed to participate. Most of them were between 25 and 45 years old, and greater age was not associated with poorer acceptability. Most of the practitioners regularly used new technologies, and 95% (20/21) connected daily to the internet, with 70% (15/21) connecting 3 times a day or more. However, only 57% (12/21) reported feeling comfortable with computers. Of the clinicians, 86% (18/21) would recommend the tool to their colleagues and 67% (14/21) stated that they would be interested in daily use of the app. Most of the clinicians (16/21, 76%) found the interface easy to use and useful. However, several clinicians noted the lack of readability (8/21, 38%) and the need to improve ergonometric features (4/21, 19%), in particular to facilitate browsing through various subsections. Some participants (5/21, 24%) were concerned about the storage of medical data and most of them (11/21, 52%) seemed to be uncomfortable with this.

Conclusions: We describe the first step of the development of a Web app combining an electronic health record and ecological momentary assessment features. This online tool offers the possibility to assess patients and to integrate medical data easily into face-to-face conditions. The acceptability of this app supports the feasibility of its broader implementation. This app could help to standardize assessment and to build up a strong database. Used in conjunction with robust data mining analytic techniques, such a database would allow exploration of risk factors, patterns of symptom evolution, and identification of distinct risk subgroups.

(J Med Internet Res 2019;21(4):e10111) doi:10.2196/10111

KEYWORDS

acceptability; feasibility studies; mobile applications; ecological momentary assessment; decision support systems, clinical; internet; outpatients; young adult; prodromal symptoms; mental health

Introduction

Optimizing Data Collection

Over the last decade, the field of medicine has evolved toward greater digitization of data in order to improve coordination and continuity of care [1,2]. Information and communication technologies have brought computer science into medical units; medical data collection software is now abundant, and its use is expanding into a variety of settings [2,3]. However, many clinicians view this new reliance on technology as an added burden to their existing workload [2,4-6]. Despite some attempts to standardize practices, no single tool has been widely incorporated into routine use [4,7].

Data collection tools are often based on information stored as uncoded free text. The diversity of practices and theoretical orientations (interindividual variability) of each practitioner brings a high variability that could be a barrier to the processing of such data for clinical decision-making purposes and research purposes [4]. Intraindividual variability is affected by many factors, such as time, personal subjective factors [4], patient interaction, consultation environment, and symptoms presented by the patient [8], and this variability is greatly increased when there is no data collection frame [9]. Several studies have shown that using standardized semistructured interviews can improve the quality and completeness of data collection [9,10]. In addition, use of standardized semistructured interviews also facilitates the collection and use of high-quality data for research purposes without increasing the clinical workload, facilitating prospective observational studies and clinical studies [11].

Thus, it is important to optimize data collection as part of routine clinical practice and decision making, to standardize the collection of a common minimum dataset, and to develop electronic data collection platforms to support these activities [12].

Objective

Our early mental disorder detection program aims to identify emerging psychiatric disorders in young outpatients. A standardized assessment is performed by a member of the clinical team to detect individual risk factors for a psychiatric disorder and to inform care. The objective of this study was to assess clinician acceptability of using a computer clinical data collection tool during a consultation. Secondary objectives of this study were to explore (1) the technical feasibility of using a connected tool during a consultation (accessibility, compatibility with the tools available in the university hospital), (2) the effectiveness of a computer interface for data collection (ease of use, ergonomics), and (3) the subjective experience of caregivers using the electronic health (eHealth) platform.

Methods

MEmind App Description

In partnership with the University of Madrid, we developed a Web app (MEmind) to allow the collection of clinical data in real time [13]. The tool allows clinicians to enter details into an electronic health record during a psychiatric evaluation, including sociodemographic details, clinical examinations, diagnoses, therapeutic factors, psychometric scales, and free-text observations.

The interface has been developed for cross-platform use on desktop computers, tablets, and smartphones, allowing use by health care professionals in different places in the practitioner's workplace. Access is restricted by a password issued to the professional. The app can be customized to ambulatory practices and different mental health research protocols, and a wide range of relevant scales can be included in routine evaluations at the practitioner's discretion.

For this study, we developed a semistructured interview script based on the standard evaluation form that the early detection team had previously used. The interview explores the patient's sociodemographic background, the patient's history, and the histories of their family members.

MEmind is also built to enable the patient to connect to another component, the personal health record, which they can access via a computer or smartphone to enter data in an ecological momentary assessment (EMA) view. Indeed, clinical assessment in psychiatry is usually based on findings from brief, regularly scheduled, in-person appointments. Although critically important, this approach reduces assessment to cross-sectional observations that miss essential information about disease course and are subject to recall bias. EMA involves repeated sampling of a person's behaviors and experiences in real time, in their natural environment. Patient self-monitoring can rely on EMA procedures and lead to participatory medicine [11]. EMA has been successfully used for real-time self-reporting of symptoms and behavior in patients with anxiety disorders or suicide ideation. Given that psychiatry clinicians have previously relied exclusively on clinical interviews for diagnosis and treatment, the field could deeply benefit from this new source of data collected in real time covering information about the patient's health state between visits. Mobile phones are generally kept on at all times and carried everywhere, making them an ideal platform for the broad implementation of EMA technology. EMA has a number of advantages. First, it gives the clinician insight into the contingencies of experience and mental states, based on prospective data. Second, EMA observations have, contrary to clinical interactions, ecological validity, reflecting real-life variation in response to real-life challenges [14]. These data can also be compared with the data collected during in-person visits and stored in regular eHealth reports or a Web-based data collection tool. We did not explore the use of EMA in this study, which focused on the clinicians' experiences.

Study Design

This study was a prospective acceptability study, designed to assess clinicians' acceptability of the interface in routine clinical use. We asked clinicians from the 4 community psychiatry departments within Brest University Medical Hospital (France) to participate between June 31 and August 31, 2015. All clinicians working with adolescents and young adults were invited to participate by email or phone and gave their signed consent before taking part in this study. The only exclusion criteria were the clinician's inability to use the computer tool and declining to participate.

After clinicians provided consent and were introduced to the study by an investigator, they were shown a video tutorial [15] to describe the functioning of the Web app. We then presented a video showing a first consultation with a patient presenting with emerging psychiatric concerns. This video reproduced the context of a consultation with a new patient, with their consent for the video to be used for this research purpose [16,17]. We asked clinicians to collect clinical data directly using the semistructured interview with the connected tool after or during the viewing. For this study, no patient clinical data were stored. The patient who appeared on the video agreed to having his image used for educational and research purposes.

We then evaluated the acceptability and technical feasibility of the app with an anonymized paper questionnaire (Figure 1).

We explored technical feasibility at each stage of the process. The questions focused on ergonomic, technical, and ethical issues, and the app's possible use in routine practice. We measured acceptability using a 10-item paper self-questionnaire (Multimedia Appendix 1 shows the French version and Multimedia Appendix 2 shows an English translation) based on the Acceptability E-scale [18], which is a generic and validated questionnaire that can accurately evaluate satisfaction with a broad range of eHealth systems. This scale has been validated in French [19]. Participants completed the questionnaire after using the Web tool. The questions covered the technical feasibility and practical use of the tool, with each item graded using a 5-point Likert scale [20,21]. Additional questions centered on participants' thoughts about daily use of computerized tools and their own habits. They were also asked to provide sociodemographic information about their age and sex. Depending on the question, other answers were given as either "yes" or "no," as a choice between various options, or as free text [22].

Figure 1. View of the connected tool.

We assessed the acceptability of the interface using a question from the Net Promoter Score, “Would you recommend this tool to your colleagues ?” This question is conventionally used in satisfaction studies and is a validated tool for this type of study [21].

Questionnaire response was anonymous. Participants submitted their data to a nonparticipating research nurse to preserve anonymity and reduce the risk of inadvertent breach of confidentiality. We digitized the answers of the self-questionnaire using double data entry to avoid transcription errors before statistical analysis. The results were described as percentage positive response.

Statistical Analysis

We tested associations between categorical variables using a chi-square test of independence on the contingency tables and exact Fisher tests. *P* values are reported throughout the text when needed. We performed factorial analysis through a multiple correspondence analysis. This method is suitable for survey analysis, as it projects an initial set of qualitative variables in a factorial space where each dimension is a combination of the initial variables. We assessed the association between the initial variables (ie, the survey questions and the answer modalities) and the reduced dimensions with hypothesis testing.

The study was conducted according to French data processing requirements and the World Medical Association Declaration of Helsinki [23]. This study received a favorable ethical opinion from the Brest Medical University Hospital Ethics Committee.

Results

Participants’ Use of New Technologies

We invited 24 clinicians, and 21 (88%) agreed to participate. Of these, 12 (57%) were between 25 and 45 years old. Table 1 presents the characteristics of the participants’ use of new technologies.

Most practitioners regularly used new technologies, 76% (16/21) used a smartphone, and 95% (20/21) connected daily to the internet, 70% (15/21) at least 3 times a day. However, only 57% (12/21) felt that they were proficient with computers.

Participants’ Views on the App

A total of 18 (86%) of the clinicians stated that they would recommend the Web app to their colleagues for data collection purposes. We found that 67% (14/21) felt that routine use in specialized assessments would be of interest, but when the subgroup analysis excluded practitioners who were only “a little” or “not at all” familiar with the computer tool, the acceptability of the software was 86% (18/21). Through a subgroup analysis, the acceptability of the tool thus increased significantly. We also performed a subgroup analysis according to the age of the participants (>55 years vs ≤55 years). An increase in age was not correlated with a decrease in acceptability of the interface (*P*=0.47).

Most clinicians (16/21, 76%) found the interface easy to use and useful. Table 2 shows data on the use of the tool. The enthusiasm for the computer tool didn’t depend on the age range of the clinicians. Although most participants had a positive experience with the platform, 2 of them (10%) found it not very useful, especially because 1 of them had encountered difficulties in connecting.

Table 1. Participants’ use of new technologies.

Characteristic	n (%)
Frequency of internet use	
Less than once a week	0 (0)
Between 5 and 10 times a week	1 (5)
Once a day	5 (25)
3 times a day	2 (10)
More than 3 times a day	12 (60)
Internet familiarity	
Not familiar at all	1 (5)
A little familiar	2 (10)
Moderately familiar	6 (30)
Quite familiar	10 (50)
Very familiar	1 (5)

Table 2. Participants' views on use of the MEmind Web app.

Response	Question item category, n (%)					
	Ease of use	Usability	Understandability	Time to complete data entry	Completeness	Usefulness
Very satisfied	13 (62)	5 (24)	3 (14)	2 (10)	2 (10)	5 (24)
Somewhat satisfied	5 (24)	12 (57)	10 (48)	14 (67)	13 (61)	11 (52)
Very satisfied or somewhat satisfied	18 (86)	17 (81)	13 (62)	16 (76)	15 (71)	16 (76)
No opinion	2 (10)	1 (5)	3 (14)	1 (5)	1 (5)	2 (10)
Somewhat dissatisfied	1 (5)	3 (14)	5 (24)	2 (10)	5 (24)	2 (10)
Very dissatisfied	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Several clinicians noted the lack of readability (8/21, 38%) and the need to improve the ergonomics (4/21, 19%) of the navigation interface within the various submenus. Several improvements were proposed to facilitate navigation and improve the ergonomics in the light of the free comments of some practitioners. Clinicians suggested several options: distributed tabs, drop-down menus, and a more compact presentation.

In general, practitioners appreciated the tool and considered that connection was easy (18/21, 86%), that usability was good (17/21, 81%), and that completing data entry was not time consuming (16/21, 76%). Chi-square tests confirmed the associations between participants' level of satisfaction and the tested criteria (*P*=.01).

In the free observations that the clinicians could provide in the questionnaire, we noted that many of them showed an interest in this new tool. They pointed out that the app made it possible to standardize the collection of certain fundamental data by providing a framework that was sufficiently flexible. Others noted the lack of completeness and suggested enhancing the interface with additional scales and complementary articles. A major point of interest was that the interface provided secure access to patient data on different sites, reduced the redundancy of examinations, and made the best use of the very large amount of data collected during consultations.

Participants' Views on Online Patient Data Storage

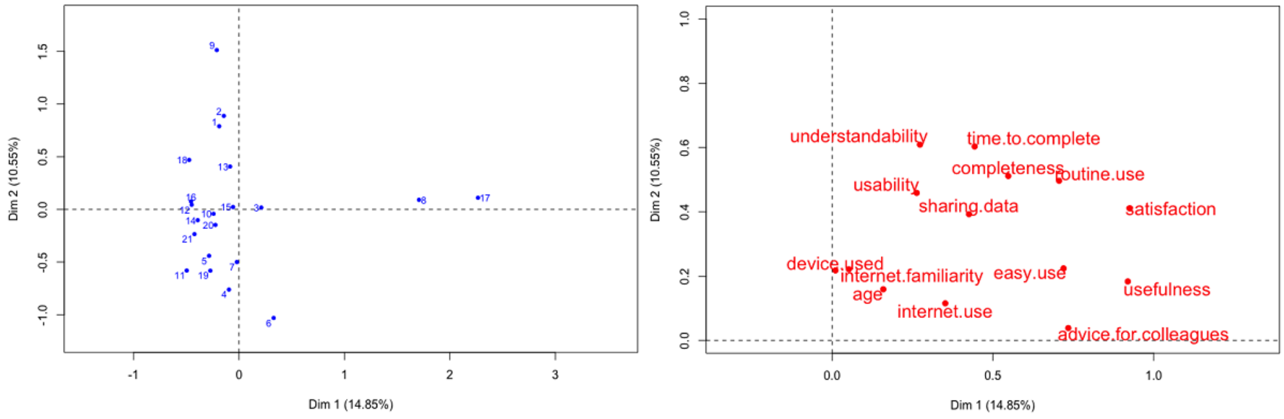
Regarding the storage of patient information on the internet, Table 3 presents the distribution of the opinions of the psychiatric clinicians who participated in the study. This was the only question about which several clinicians did not wish to express an opinion. Together with those who selected the option "no opinion," 52% (11/21) did not give any opinion on the storage of patients' medical information on the internet. This is consistent with previous reports of physician concerns with online medical data storage [3].

A multiple correspondence analysis gave a synthetic view of the answers and the links between the satisfaction criteria. Figure 2 shows a 2-dimensional representation of the answers. The proximity between the points reflects the association between the survey answers. The results highlight some relevant trends: satisfaction was strongly linked to the tool completeness and a possible routine use, as well as ease of use and usefulness (dimension 1 on the x-axis). Understandability, usability, and time to complete data entry are also grouped as common criteria, expressing the ergonomics of the tool. Participants' experience with the internet and use of devices to connect to the internet were also relevant factors for the acceptability of the tool.

Table 3. Practitioners' views on online data storage, in response to the question "What do you think about storing patients' medical data on the internet?"

Response	n (%)
Very dissatisfied	2 (10)
Somewhat dissatisfied	3 (14)
No opinion	7 (33)
Somewhat satisfied	4 (19)
Very satisfied	1 (5)
No answer	4 (19)

Figure 2. Multiple correspondence analysis. Projection of the variables onto the factorial map (left) and projection of the clinicians' answers (right). Only the 2 first dimensions (Dim 1 and Dim 2) are represented.



Discussion

Principal Findings

Our data support the feasibility of incorporating electronic health record data collection tools into routine clinical practice, to support the implementation of standardized data collection in outpatient services. The acceptability of a Web app for systematic clinical data collection was good among the clinicians surveyed. Specifically, more than 86% of clinicians would recommend the tool to their colleagues, 76% found it useful, and 67% would consider its routine use. Although only 57% of participants felt comfortable with computer use, 70% of them used it on a daily basis, and the acceptability of the Web app was very good. Importantly, this seems to reflect confidence in the tool and interest in these new ways of collecting information. There was a strong interest in this new type of interface, but there were also many issues related to medical confidentiality, and there was still some skepticism about online storage of patient data.

Limitations

The participation rate in this study was 88%, and this sample was representative of the clinicians working in psychiatry at the Brest Medical University Hospital. However, the acceptability study we conducted had several limitations. The number of participants was small, and replication of the results on a larger scale in an everyday practice would be beneficial. Moreover, the semistructured interview used for this study followed the consultation model used for evaluations but did not contain psychometric scales or new items. Some participants who were more enthusiastic about the tool proposed adding new data to collect to the interface. It would be interesting to assess the acceptability of a tool that would change practices and the data collected, for example, by adding clinical scales or items that are not usually sought after. Indeed, clinicians are often reluctant to adopt tools that change the practices they have built over the years [3]. It seems, however, that the possibility of adding free text allowed clinicians to feel that they retained control over the tool, a concern that had already been identified in the literature [6].

The only elements that affected acceptability were technical aspects concerning ergonomics or difficulties of use related to connection problems at the time of testing. Age did not seem to have any influence on acceptability, whereas familiarity with computer tools did.

Several criticisms were stated regarding the readability of the interface. These anomalies did not prevent the use of the interface, which was simple but sometimes confused the user. Nevertheless, technical changes have to be made to correct display problems (tab alignments, page resizing. and title placement) and to make navigation within the interface more fluid, in particular by adding drop-down menus. These data are consistent with recent results showing the influence of ergonomic aspects on user experience and acceptability [17].

The study of the routine use of the interface would complete these results. It could also be interesting to test the use of such an interface by patients in their living environment. This may offer a new way of dealing with the symptoms they are reporting [13,24]. It is possible that future medical data tools may allow data to be collected during consultations and also on an ongoing basis in the patient's living environment, using a combination of electronic medical records and EMA. This would allow clinicians to monitor the evolution of symptoms in their naturalistic setting [25,26]. The use of standardized questionnaires allows for homogeneous collection of clinical data for an informed medical decision. This also makes it possible to consider the use of health record databases for research or collaborations.

Data Processing, Interest, and Ethical Issues

The growing use of tools dedicated to the collection of medical data, computerized medical health records, and communication tools is raising many ethical issues concerning confidentiality. This concern was reflected in the responses given by participants. However, this was the only question that more than one-third of clinicians did not answer, which exacerbates this concern.

Medical health data are being computerized within a cultural context of mistrust toward new technologies. The great ease of interpersonal communication and the flow of information entails a feeling of insecurity among individuals regarding confidentiality of these data [27]. Indeed, the use of computer tools makes it possible to heighten mistrust of transmission of information and the fear that personal data could be stolen [28].

The history of computer communication systems shows that, despite precautions taken by program developers, it is difficult to prevent data from being intercepted [29]. This is the most challenging question for clinicians: 33% of our participants did not express an opinion on the storage of data on the internet. If the participants who did not answer this question are included, the percentage increases to 42%. However, those who expressed an opinion were equally divided between satisfied (24%; very satisfied 5%, somewhat satisfied 19%) and dissatisfied (24%; somewhat dissatisfied 14%, very dissatisfied 10%). These results reflect a certain caution among the clinicians who did not seem to be opposed to the use of connected tools, but may have been concerned about engaging in a process they did not understand well.

There is thus a risk of revealing confidential information through cross-checking of data from various computer files and a risk of infringement of individual freedoms to the benefit of certain organizations, in particular administrative, financial, or insurance companies [29-31]. On the other hand, information technology tools are not always reliable in their handling (complexity of procedures, equipment breakdown or theft, loss or alteration of files, viruses, etc). Computer security can create anxiety for both the professional and the patient, particularly with respect to privacy issues. We found this concern among clinicians who took part in this acceptability study.

It is in the public interest to preserve citizens' confidence in the confidentiality of the health care system. Particular attention must be paid to combining the use of information technology tools with the care of patients. In France, this concern is reflected in the legislative provisions of the law concerning the secrecy of health information and health data hosts [32-34]. These texts state that electronic medical records contain medical information and as such should be regarded as private and confidential. It also seems necessary to question the ownership of and access to the data collected. Do the data belong to the patient, the state, or the company that hosts the data or develops the interface? These questions also raise the possibility of using data collected in daily practice for research without explicit patient consent. The security of computerized personal health information systems is therefore an ethical imperative.

Conclusion

We describe the first step of the development of a Web app combining an electronic health record and EMA features. This online tool offers the possibility of assessing patients and integrating medical data easily into face-to-face situations. The acceptability of this app supports the feasibility of its broader implementation. This app could help to standardize assessment and to build up a strong database. Used in conjunction with robust data mining analytic techniques, such a database would allow exploration of risk factors, patterns of symptom evolution, and identification of distinct risk subgroups.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire (French).

[PDF File (Adobe PDF File), 83KB - jmir_v21i4e10111_app1.pdf]

Multimedia Appendix 2

Questionnaire (English).

[PDF File (Adobe PDF File), 85KB - jmir_v21i4e10111_app2.pdf]

References

1. Adler-Milstein J, Everson J, Lee SD. EHR adoption and hospital performance: time-related effects. Health Serv Res 2015 Dec;50(6):1751-1771 [FREE Full text] [doi: 10.1111/1475-6773.12406] [Medline: 26473506]
2. Kaipio J, Lääveri T, Hyppönen H, Vainiomäki S, Reponen J, Kushniruk A, et al. Usability problems do not heal by themselves: national survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Jan;97:266-281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
3. Bruns EJ, Hook AN, Parker EM, Esposito I, Sather A, Parigoris RM, et al. Impact of a web-based electronic health record on behavioral health service delivery for children and adolescents: randomized controlled trial. J Med Internet Res 2018 Jun 14;20(6):e10197 [FREE Full text] [doi: 10.2196/10197] [Medline: 29903701]
4. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]
5. Jamoom EW, Heisey-Grove D, Yang N, Scanlon P. Physician opinions about EHR use by EHR experience and by whether the practice had optimized its EHR use. J Health Med Inform 2016 Jul 30;7(4):1000240 [FREE Full text] [doi: 10.4172/2157-7420.1000240] [Medline: 27800279]

6. Archer N, Cocosila M. A comparison of physician pre-adoption and adoption views on electronic health records in Canadian medical practices. J Med Internet Res 2011 Aug 12;13(3):e57 [FREE Full text] [doi: 10.2196/jmir.1726] [Medline: 21840835]
7. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]
8. Kazmi Z. Effects of exam room EHR use on doctor-patient communication: a systematic literature review. Inform Prim Care 2013;21(1):30-39 [FREE Full text] [doi: 10.14236/jhi.v21i1.37] [Medline: 24629654]
9. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. J Intern Med 2013 Dec;274(6):547-560. [doi: 10.1111/joim.12119] [Medline: 23952476]
10. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-173 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.006] [Medline: 25463966]
11. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: transition to participatory and personalized medicine in mental health. J Med Internet Res 2018 Jan 03;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
12. Addington J, Cadenhead KS, Cannon TD, Cornblatt B, McGlashan TH, Perkins DO, North American Prodrome Longitudinal Study. North American Prodrome Longitudinal Study: a collaborative multisite approach to prodromal schizophrenia research. Schizophr Bull 2007 May;33(3):665-672 [FREE Full text] [doi: 10.1093/schbul/sbl075] [Medline: 17255119]
13. Berrouiguet S, Barrigón ML, Brandt SA, Ovejero-García S, Álvarez-García R, Carballo JJ, MEmind Study Group, et al. Development of a web-based clinical decision support system for drug prescription: non-interventional naturalistic description of the antipsychotic prescription patterns in 4345 outpatients and future applications. PLoS One 2016;11(10):e0163796 [FREE Full text] [doi: 10.1371/journal.pone.0163796] [Medline: 27764107]
14. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Dec;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
15. Tutoriel MEmind français.: MEmind; 201 Oct 07. URL: https://www.youtube.com/watch?v=eLTiERqjEoA [accessed 2019-01-17] [WebCite Cache ID 75UqyQNrh]
16. Thomas KW, Dayton CS, Peterson MW. Evaluation of internet-based clinical decision support systems. J Med Internet Res 1999;1(2):E6 [FREE Full text] [doi: 10.2196/jmir.1.2.e6] [Medline: 11720915]
17. Duman-Lubberding S, van Uden-Kraan CF, Peek N, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. An eHealth application in head and neck cancer survivorship care: health care professionals' perspectives. J Med Internet Res 2015 Oct 21;17(10):e235 [FREE Full text] [doi: 10.2196/jmir.4870] [Medline: 26489918]
18. Tariman JD, Berry DL, Halpenny B, Wolpin S, Schepp K. Validation and testing of the Acceptability E-scale for web-based patient-reported outcomes in cancer care. Appl Nurs Res 2011 Feb;24(1):53-58 [FREE Full text] [doi: 10.1016/j.apnr.2009.04.003] [Medline: 20974066]
19. Micoulaud-Franchi J, Sauteraud A, Olive J, Sagaspe P, Bioulac S, Philip P. Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. Psychiatry Res 2016 Mar 30;237:196-200. [doi: 10.1016/j.psychres.2016.01.043] [Medline: 26809367]
20. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
21. Singh N, Hess E, Guo G, Sharp A, Huang B, Breslin M, et al. Tablet-based patient-centered decision support for minor head injury in the emergency department: pilot study. JMIR Mhealth Uhealth 2017 Sep 28;5(9):e144 [FREE Full text] [doi: 10.2196/mhealth.8732] [Medline: 28958987]
22. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
23. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bull World Health Organ 2001;79(4):373-374 [FREE Full text] [Medline: 11357217]
24. Berrouiguet S, Courtet P, Larsen ME, Walter M, Vaiva G. Suicide prevention: towards integrative, innovative and individualized brief contact interventions. Eur Psychiatry 2018 Jan;47:25-26. [doi: 10.1016/j.eurpsy.2017.09.006] [Medline: 29096129]
25. Berrouiguet S, Barrigón ML, Brandt SA, Nitzburg GC, Ovejero S, Alvarez-Garcia R, et al. Ecological assessment of clinicians' antipsychotic prescription habits in psychiatric inpatients: a novel web-and mobile phone–based prototype for a dynamic clinical decision support system. J Med Internet Res 2017 Jan 26;19(1):e25 [FREE Full text] [doi: 10.2196/jmir.5954] [Medline: 28126703]
26. Barrigón ML, Berrouiguet S, Carballo JJ, Bonal-Giménez C, Fernández-Navarro P, Pfang B, MEmind study group. User profiles of an electronic mental health tool for ecological momentary assessment: MEmind. Int J Methods Psychiatr Res 2017 Mar;26(1). [doi: 10.1002/mpr.1554] [Medline: 28276176]

27. El Emam K, Neri E, Jonker E, Sokolova M, Peyton L, Neisa A, et al. The inadvertent disclosure of personal health information through peer-to-peer file sharing programs. J Am Med Inform Assoc 2010;17(2):148-158 [FREE Full text] [doi: 10.1136/jamia.2009.000232] [Medline: 20190057]
28. El Emam K, Neri E, Jonker E. An evaluation of personal health information remnants in second-hand personal computer disk drives. J Med Internet Res 2007 Sep 30;9(3):e24 [FREE Full text] [doi: 10.2196/jmir.9.3.e24] [Medline: 17942386]
29. Woodward B. The computer-based patient record and confidentiality. N Engl J Med 1995 Nov 23;333(21):1419-1422. [doi: 10.1056/NEJM199511233332112] [Medline: 7477127]
30. Schoenberg R, Safran C. Internet based repository of medical records that retains patient confidentiality. BMJ 2000 Nov 11;321(7270):1199-1203 [FREE Full text] [Medline: 11073513]
31. Corlosquet-Habart M, Janssen J. Big Data for Insurance Companies. Volume 1. Hoboken, NJ: John Wiley & Sons; 2018.
32. Sittig DF, Singh H. Legal, ethical, and financial dilemmas in electronic health record adoption and use. Pediatrics 2011 Apr;127(4):e1042-e1047 [FREE Full text] [doi: 10.1542/peds.2010-2184] [Medline: 21422090]
33. Code de la santé publique. Article R1111-9 à R1111-15.: Légifrance; 2018 Apr 1. URL: https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006072665&idArticle=LEGIARTI000006908146&dateTexte=&categorieLien=cid
34. Loi no 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé.: Légifrance; 2019 Jan 13. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000227015 [accessed 2019-01-17] [WebCite Cache ID 75UqsdAul]

Abbreviations

eHealth: electronic health
EMA: ecological momentary assessment

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by T Rashid Soron, M Bujnowska-Fedak; comments to author 01.04.18; revised version received 27.08.18; accepted 11.10.18; published 25.04.19.

Please cite as:
Lemey C, Larsen ME, Devylder J, Courtet P, Billot R, Lenca P, Walter M, Baca-García E, Berrouiguet S
Clinicians' Concerns About Mobile Ecological Momentary Assessment Tools Designed for Emerging Psychiatric Problems: Prospective Acceptability Assessment of the MEmind App
J Med Internet Res 2019;21(4):e10111
URL: https://www.jmir.org/2019/4/e10111/
doi:10.2196/10111
PMID:31021327

Corrigenda and Addenda

Correction: Automatic Extraction of Mental Health Disorders From Domestic Violence Police Narratives: Text Mining Study

George Karystianis[1], BSc, MSc, PhD; Armita Adily[1], BMedSc, MPH, PhD; Peter Schofield[2], BSc (Hons), BS, Med, MSc, MD; Lee Knight[1], MForMentH, MN; Clara Galdon[3], BA; David Greenberg[4], MB, ChB, FFPsych (SA), MMed, FRCPC, FRANZCP; Louisa Jorm[5], BVSc, MSc, PhD; Goran Nenadic[6], BSc, MSc, PhD; Tony Butler[1], BSc, MSc, PhD

[1]Kirby Institute, Faculty of Medicine, University of New South Wales, Sydney, Australia
[2]Neuropsychiatry Service, Hunter New England Health, Newcastle, Australia
[3]Victims Services, New South Wales Department of Justice, Sydney, Australia
[4]School of Psychiatry, University of New South Wales, Sydney, Australia
[5]Centre for Big Data Research in Health, University of New South Wales, Sydney, Australia
[6]School of Computer Science, University of Manchester, Manchester, United Kingdom

Corresponding Author:
George Karystianis, BSc, MSc, PhD
Kirby Institute
Faculty of Medicine
University of New South Wales
Level 6, Wallace Wurth Building High Street
Kensington NSW
Sydney, 2052
Australia
Phone: 61 (2) 9385 0900
Email: gkarystianis@kirby.unsw.edu.au

Related Article:

Correction of: http://www.jmir.org/2018/9/e11548/

(J Med Internet Res 2019;21(4):e13007) doi:10.2196/13007

The authors of “Automatic Extraction of Mental Health Disorders From Domestic Violence Police Narratives: Text Mining Study” (J Med Internet Res 2018;20(9):e11548) incorrectly under-reported the number of ICD-10 level 1, 2 and 3 categories used for mapping the identified mental health disorder mentions and the number of domestic violence events at certain mental health disorders in the Abstract and in the Results section. Note that the acronym “POI(s)” is frequently used within the text to mean “person(s) of interest”; this use has been preserved in quoted sections of text below.

In the Abstract, the authors have incorrectly listed the number of domestic violence events with depression in victims as “22.30%; 3258” instead of “22.25%, 3269” and in POIs as “18.73%, 8918” instead of “18.70%, 8944”. Additionally, the alcohol abuse domestic violence event percentage for POIs were incorrectly listed as “12.24%, 5829” instead of “12.19%, 5829” and the percentage and number of domestic violence events for various anxiety disorders for victims were incorrectly listed as “11.43%, 1671” instead of “11.66%, 1714”. All aforementioned errors have been corrected.

In the “Data” subsection of the Methods, the authors have listed a total of 24 categories instead of 26 in Textbox 1. The authors with to add two more categories that were omitted: “25. Unspecified diseases of the nervous system: eg, neurological disorder” and “26. Systemic atrophies primarily affecting the central nervous system: eg, Huntington disease”. The authors wish to remove the example “Huntington disease” from the 15 category level 1 “Diseases of the nervous system: eg, Alzheimer disease, Huntington disease, frontotemporal dementia”. Also, the authors did not write down correctly the name of the category 15 and they wish to change its name from “Diseases of the nervous system” to “Other degenerate diseases of the nervous system”.

In the “Mapping of Extracted Mental Health Disorder Mentions to the International Classification of Diseases, Tenth Revision” subsection of the Methods, the authors incorrectly stated that “The first level was the most generic (24 categories)” instead of “26 categories”; “The second and third levels of mapping had 71 and 163 categories, respectively” instead of “62 and 98 categories”; and “a fourth level of *ICD-10* classification (25 categories) was recorded” instead of “a fourth level of *ICD-10* classification (27 categories)”. All aforementioned errors have been corrected.

The authors have incorrectly reported the number of certain mental health disorder mentions at the first, second and third levels in the “Large-Scale Corpus Application” subsection of the Results. In Table 4, the authors listed the number of mental health disorder mentions in the “POIs only” row as “25,211” instead of “21,127” at the third level; “47,600” instead of “47,831” at the second level; and “79,727” instead of “81,942” at the first level. The authors incorrectly reported that the number of mental health disorder mentions for the “Victims only” as “14,609” instead of “14,695” at the second level and as “20,774” instead of “21,290” at the first level. The authors incorrectly reported the number of mental health disorder mentions in the “Total” row as “62,209” instead of “62,526“ at the second level and “100,501” instead of “103,232” at the first level. All aforementioned errors have been corrected.

Subsequently, the same errors were listed in the third paragraph of the “Large-Scale Corpus Application” subsection which mentions Table 4, and the number of classified mental health disorder mentions have changed to reflect this:

> *The total number of classified mental health disorder mentions at the first level was 103,232, whereas 62,526 mental health disorder mentions contained sufficient information allowing them to be mapped to the second level, with one-third of mentions (32,479, 31.46%) mapped to the third level (Table 4).*

The number of mentions for persons of interest and victims in Tables 5 and 6 for certain mental health disorders have been listed incorrectly by the authors.

In Table 5, “Diseases of the nervous system” has been changed to “Other degenerate diseases of the nervous system”. The authors wish to change the number of mentions in the “POIs” column for the following mental health disorders:

- The value for “Mood (affective) disorders” has been changed from “14,566” to “15,330”
- The value for “Behavioral and emotional disorders with onset usually occurring in childhood and adolescence” has been changed from “9034” to “9848”
- The value for “Anxiety, dissociative, stress-related, somatoform, and other nonpsychotic mental disorders” has been changed from “3518” to “3755”
- The value for “Mental and behavioral disorders due to psychoactive substance use” has been changed from “6762” to “6790”
- The value for “Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders” has been changed from “5598” to “5771”
- The value for “Intellectual disability” has been changed from “1444” to “1517”
- The value for “Mental disorders due to known physiological conditions” has been changed from “557” to “559”
- The value for “Pervasive and specific developmental disorders” has been changed from “1686” to “1775”
- The value for “Disorders of adult personality and behavior” has been changed from “1308” to “1340”
- The value for “Other degenerate diseases of the nervous system” has been changed from “78” to “62”
- The value for “Behavioral syndromes associated with physiological disturbances and physical factors” has been changed from “29” to “31”

Also in Table 5, the authors wish to change the number of mentions in the “Victims” column for the following mental health disorders:

- The value for “Mood (affective) disorders” has been changed from “4734” to “4946”
- The value for “Behavioral and emotional disorders with onset usually occurring in childhood and adolescence” has been changed from “2143” to “2224”
- The value for “Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders” has been changed from “2113” to “2261”
- The value for “Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders” has been changed from “1020” to “1032”
- The value for “Intellectual disability” has been changed from “907” to “939”
- The value for “Mental disorders due to known physiological conditions” has been changed from from “648” to “649”
- The value for “Disorders of adult personality and behavior” has been changed from “406” to “420”
- The value for “Other degenerate diseases of the nervous system” has been changed from “61” to “52”

The authors wish to add two new rows to Table 5:

- “Systematic atrophies primarily affecting the central nervous system” (POIs=11; victims=6)
- “Unspecified diseases of the nervous system” (POIs=6; victims=3)

Consequently, the same errors were listed in the fourth paragraph of the “Large-Scale Corpus Application” subsection which mentions Table 5, and the number of classified mental health disorder mentions have been changed to reflect this:

> *At the first level (Table 5), almost one-third of the 81,942 mentions of mental health disorders (32.46%, 26,598) for the POI and one-fifth (22.79%, 4851) for victims had “unspecified mental health disorders” not explicitly recorded in the narratives by the attending police officer(s). “Mood (affective) disorders” (eg, bipolar disorder, depression) had the highest number of mentions among POIs (15,330, 18.71%) and victims (4946, 23.23%) with “mental and behavioral disorders due to psychoactive substance use” (including alcohol abuse) ranking fourth and fifth for both POIs (6790, 8.29%) and victims (1259, 5.91%), respectively. In all, 12.02% of POIs (9848) and 10.45% of victims (2224) had mentions of “behavioral and emotional disorders with their onset usually occurring in childhood and adolescence” (eg, “attention deficit hyperactivity disorders,” “conduct disorders”) being the third and*

> *fourth biggest group of disorders in both POIs and victims. Although mentions of "intellectual disabilities" among POIs (1517, 1.85%) were higher in number than in the victims (939, 4.41%), the rates were higher among victims than POIs. Mentions of traumatic brain injury (eg, "the victim has suffered a brain injury due to a car accident") were reported for 0.84% of POIs and 1.17% victims (688 and 250 mentions, respectively).*

In Table 6, the authors wish to change the number of mentions in the "POIs" column for the following mental health disorders:

- The value for "Major depressive disorder, single episode" has been changed from "8918" to "8944"
- The value for "Bipolar disorder" has been changed from "5448" to "5449"
- The value for "Other behavioral and emotional disorders with onset usually occurring in childhood and adolescence" has been changed from "4888" to "4880"
- The value for "Other anxiety disorders" has been changed from "2420" to "2446"
- The value for ”Pervasive developmental disorder” has been changed from "1635" to "1721"
- The value for "Specific personality disorders" has been changed from "1279" to "1310"
- The value for "Conduct disorders" has been changed from "877" to "903"
- The value for "Reaction to severe stress, and adjustment disorders" from "787" to "790".

Also in Table 6, the authors wish to replace the number of mentions in the "Victims" column for the following mental health disorders:

- The value for "Major depressive disorder, single episode" has been changed from "3258" to "3269"
- The value for "Schizophrenia" has been changed from "846" to "849"
- The value for "Other anxiety disorders" has been changed from "1671" to "1714"
- The value for "Pervasive developmental disorder" has been changed from "462" to "477"
- The value for "Specific personality disorders" has been changed from "358" to "372"
- The value for "Conduct disorders" has been changed from "119" to "121"

Consequently, the same errors were listed in the fifth paragraph of the "Large-Scale Corpus Application" subsection which mentions Table 6, and the number of mentions for mental health disorders for both persons of interest and victims have been changed to reflect this:

> *In the second level categories (Table 6), "alcohol abuse" was the second highest mental health disorder among POIs (5829, 12.19%) and the fifth highest reported among victims (1180, 8.03%) reinforcing the established link between DV and alcohol use [31-33]. Additionally, there were 644 victims with "dementia, unspecified" (4.38%, 644/14,609) and 546 POI ones (1.14%, 546/47,600).*

The authors wish to state that the requested changes do not affect the reported conclusions on the paper. Since the number of mental health disorder mentions in domestic violence events was slightly under-reported, the conclusions remain the same. The only difference is the increase of certain mentions of mental health disorders. The original statistics that were reported were marginally different in the last out of two decimals.

The correction will appear in the online version of the paper on the JMIR website on April 5, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Edited by G Eysenbach; submitted 05.12.18; this is a non–peer-reviewed article;accepted 13.12.18; published 05.04.19.

Please cite as:
Karystianis G, Adily A, Schofield P, Knight L, Galdon C, Greenberg D, Jorm L, Nenadic G, Butler T
Correction: Automatic Extraction of Mental Health Disorders From Domestic Violence Police Narratives: Text Mining Study
J Med Internet Res 2019;21(4):e13007
URL: https://www.jmir.org/2019/4/e13007/
doi:10.2196/13007
PMID:30951492

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org