Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 22 (2020), Issue 1 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Viewpoints

Reviews

Original Papers

Viewpoint

A Virtual Home for the Virtual Clinical Trial

Susan Persky[1], PhD

Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD, United States

Corresponding Author:
Susan Persky, PhD
Social and Behavioral Research Branch
National Human Genome Research Institute
31 Center Dr Bldg 31 Rm B1B36
Bethesda, MD, 20892
United States
Phone: 1 301 443 0098
Email: perskys@mail.nih.gov

Abstract

Virtual clinical trials (VCTs) can satisfy the need for rigorous clinical trials by using distributed technological solutions that eliminate the need for a physical trial site. This report explores potential benefits of using virtual reality (VR) to provide a "virtual site" for VCTs, a shared immersive hub in which VCT participants could experience elements of the trial and interact with the trial team. VR is a communication technology that has been emerging alongside the development of VCTs, although they have never been merged in a substantial way. Many of the gaps within the VCT paradigm are areas in which VR excels. VR environments are standardized and precisely uniform, the technology allows introduction of an almost endless set of stimuli to participants' visual and auditory systems, and VR systems are adept at capturing precise movement and behavioral data. Although VR has not yet found its way into VCTs, much of the groundwork for such integration has been laid through research and technological development achieved in the past few years. Future implementation of VR within VCTs could move us from site-less trials to those with a virtual site serving as a hub for trial information provision, interaction with trial representatives, administration of evaluations and assessments, and more.

(J Med Internet Res 2020;22(1):e15582) doi:10.2196/15582

KEYWORDS

virtual reality; virtual clinical trials; eHealth

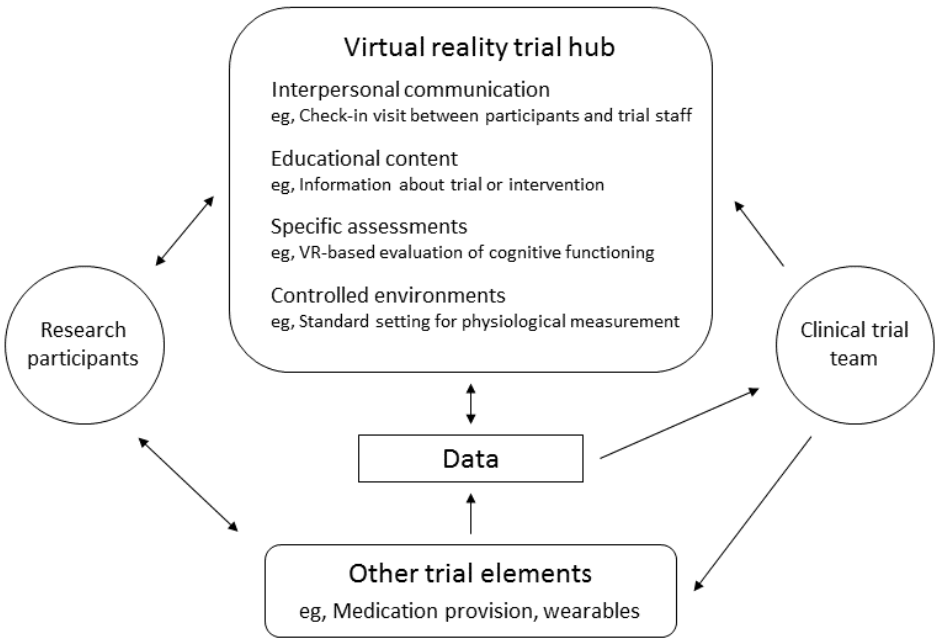
Introduction

Rapid innovation in health and health care technologies, interventions, and products require the attendant proliferation of rigorous clinical trials to evaluate their efficacy and effectiveness. To meet these needs, the field has been slowly moving toward increasing adoption of virtual clinical trials (VCTs). VCTs go by many names (eg, site-less, remote, decentralized, direct-to-patient, and patient-centered trials), but these generally involve evaluating the effect of a clinical intervention (often a pharmaceutical product) within research participants' own settings, as opposed to a clinical trial site. They satisfy the need for rigorous clinical trials performed among a diverse sampling of the appropriate population, using benefits conferred by distributed technological solutions (eg, mobile phone apps) that are often already present in the homes and routines of research participants under study. Boosters of the model submit that VCTs can reduce costs, shorten trial timelines, increase protocol adherence, and boost recruitment numbers and participant diversity, while simultaneously allowing for continuous real-world data collection in the context of real-life settings and events [1,2]. These trials typically provide participant access to research teams through Web-based portals, sometimes provide home visits, and collect data through networked wearables and medical devices, surveys, and other means. Although there are certainly complexities related to the massive amount of data generated, safety concerns, and other considerations, this approach holds great potential for growth. Indeed, if we think beyond trials involving pharmaceuticals and therapeutics, these site-less trials have been flourishing for years in the context of evaluating products and interventions (such as mobile health apps) intended to be widely deployed in free-living environments.

Still, since the dawn of the VCT, there has been argument for retaining the clinical site. Distributing data collection introduces a lack of standardization, a lack of relationship between the trial team and participants, the inability to collect certain types of data, and a lack of researcher knowledge about the contextual

details of data collection, including whether the intended individual is participating in trial activities [3]. For these and other reasons, there may be benefits to providing a "virtual site" for VCTs—a shared immersive hub in which every VCT participant could experience trial elements and interact with the trial team (Figure 1).

Figure 1. Relationships among virtual clinical trial elements incorporating virtual reality.



Virtual Reality to Augment Virtual Clinical Trials

It is time to consider whether and how virtual reality (VR) could be used to augment VCTs. VR is a communication technology that has been emerging alongside VCTs, and although they share a component of their name, they have not been merged in a substantial way. There are several arguments for doing so. Many of the gaps left by the VCT paradigm are areas in which VR excels. VR environments are digital; thus, they are very precisely uniform, and the technology allows introduction of an almost endless set of stimuli to participants' visual and auditory systems (and sometimes other sensory systems as well). VR is adept at capturing precise movement and behavior data; indeed, user movement drives the very operation of VR systems. Such data collection could also provide information on participant adherence to elements of the trial regimen, as well as data points to demonstrate that the intended individual is participating in the remote trial. Indeed, individuals' VR data are unique and identifiable in terms of elements such as eye height, gait, and movement patterns [4]. Although other related technologies, such as mobile phone apps, and nonimmersive virtual environments may likewise confer benefits for VCTs, these particular elements (immersion in uniform but realistic simulations, unique data signatures generated by body movement) are unique to VR.

Virtual Reality Research Relevant for Remote Interactions

With the aging of the American population, evaluation and therapy in home settings have become a major focus within the VR industry. For example, researchers have begun to assess the ability of patients to use VR apps for physical and occupational therapy self-assessment and exercises in the home, with generally favorable results [5-7]. This research has sensibly occurred primarily in domains where mobility can be an issue and where home-based activities are already the norm. This is a potential starting point for VR-based motion or mobility assessments in VCTs where interventions under study influence these processes. This is, however, only a single area of development. Given a very active VR development community around health, health care, and wellness, there are many use cases soon to be ripe for the picking. For example, VR assessments of neuropsychological processes and outcomes have been developed both in the context of noninterventional natural history trials and in pharmaceutical trial contexts. For example, in laboratory-based work, researchers have used VR classrooms and driving simulators to evaluate effects of psychostimulants [8-10]. Researchers have also developed VR environments to elicit and assess stress reactivity in both laboratory-standard and in ecologically realistic ways [11-13]. Cognitive and executive function evaluation is another area of

active VR development, wherein testing takes place in a lifelike virtual environment (eg, a virtual grocery store) to provide consistency between trials and between patients [14]. Elicitation and evaluation of craving and substance use behavior using VR environments have become quite sophisticated in recent years [15,16], alongside VR-based food choice measures [17-19]. These tools could provide standardized environments in which to measure specific reactions and behaviors associated with interventions under study. They also allow evaluation in the face of such stimuli without sending participants into the way of potential harm. Although many VR environments are typically designed to bring elements of the real world (bars, liquor stores, cafeterias, etc) into sterile laboratory environments, they could provide standardized environments within a home context just as compellingly.

In addition, note that applications of VR to VCTs need not be complicated or complex. Consider measuring resting blood pressure among VCT participants in the variety of home environments where this might occur—alone in a quiet room, versus surrounded by active children, watching a cooking show versus watching a true-crime show. Now consider what might be gained by having these measurements taken while all participants are relaxing on the same VR beach watching the waves roll in.

Interpersonal Factors in Virtual Clinical Trials

Also critical is the social component of clinical trials, along with the trust that can develop between participants and researchers. A VR-based communication hub for VCTs could reinsert some of the human element into these distributed studies. Many researchers have created VR clinical settings that could be leveraged as familiar and trustworthy contexts in which to convey clinical information to patients. Research on social VR, as well as work on telemedicine, has shown the ability of VR to support true social interactions and therapeutic alliance [20,21]. Use of VR to support interpersonal interaction between patients and trial staff could also be beneficial in the case of single-blind trials, wherein communication could be filtered or elements could be automated to reduce concerns about researcher expectations seeping into the encounter [22]. Although VR can convincingly mimic a clinical interaction [23,24], there may also be reason to embellish and explore new possibilities in these information exchanges. Patient education could be enhanced by bringing VCT participants into VR educational environments, for example, to help patients visualize health and medical data [25,26]. Clinical trialists could have a full arsenal of VR visualizations and demonstrations when explaining trial procedure, disease processes, medical procedures, treatment methods, and so on, within the consent process and as trials proceed.

Why Now?

Deploying VR to participant home environments has become increasingly possible because of massive growth in VR technology in recent years, as well as (much more slowly) growing ownership of VR hardware among consumers. VR hardware is available in a variety of form factors and price points. Although eventually, many trial participants may be able to bring their own VR equipment to trials, much as they might bring their own computer or mobile phone, at present, this would be an unlikely scenario. However, trial-provided equipment is common, and certainly, increased control over and standardization of trial equipment are likely of benefit at this stage.

Limitations

This is not to suggest that VR could or should be integrated into all, or even most, VCTs. There will be a great variation in the potential benefits of VR for a given trial, and these should be always weighed against potential risks and downsides. Already identified risks of VR in research include privacy and data security, as well as potential health and comfort risks of equipment use [27]. There are also some VR tools that, although effective, are best experienced while under the direct care of a health care professional, such as those aimed at addressing posttraumatic stress disorder [28]. These applications may not be appropriate for remote use. There may additionally be populations for whom home VR use will not be appropriate (eg, individuals with certain neurological conditions). Finally, as with all VCT tools, care must be taken to ensure that use protocols are easy to understand, seamless, and free of frustrations for participant populations.

Conclusions

VCTs and VR have grown up alongside but parallel to one another. The next steps toward enabling the integration of VR within VCTs include increasing distributed, site-less research on VR-based interventions. Indeed, clinical research within the VR research community itself is still somewhat emerging [29,30]. Most VR-centric health and medical research is currently performed in laboratory or medical settings; few evaluation trials have been distributed into individuals' home contexts. By taking several smaller steps, such conducting more VCTs evaluating VR-based interventions and using VR tools as part of traditional clinical trials, the frameworks and optimal approaches for engaging research participants and integrating VR into workflows can be developed. As such, we can outline the pathway moving from site-less VCTs to trials that have a virtual site, serving as a hub for trial information provision, interaction with the trial team, administration of assessments, and more.

Acknowledgments

This work was supported by the Intramural Research Program of the National Human Genome Research Institute. The author thanks Charlotte Hagerman for comments on an earlier version of this document and Emma Schopp and Sydney Telaak for assistance with figure creation.

Conflicts of Interest

None declared.

References

1. Hirsch IB, Martinez J, Dorsey ER, Finken G, Fleming A, Gropp C, et al. Incorporating site-less clinical trials into drug development: a framework for action. Clin Ther 2017 May;39(5):1064-1076 [FREE Full text] [doi: 10.1016/j.clinthera.2017.03.018] [Medline: 28413148]
2. Sommer C, Zuccolin D, Arnera V, Schmitz N, Adolfsson P, Colombo N, et al. Building clinical trials around patients: evaluation and comparison of decentralized and conventional site models in patients with low back pain. Contemp Clin Trials Commun 2018 Oct;11:120-126 [FREE Full text] [doi: 10.1016/j.conctc.2018.06.008] [Medline: 30094387]
3. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009 May 3;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
4. Kupin A, Moeller B, Jiang Y, Banerjee NK, Banerjee S. Task-driven biometric authentication of users in virtual reality (VR) environments. In: MultiMedia Modeling. Cham: Springer International Publishing; 2019:55-67.
5. Cui J, Yeh S, Lee S. Wearable sensors integrated with virtual reality: a self-guided healthcare system measuring shoulder joint mobility for frozen shoulder. J Healthc Eng 2019;2019:7681237 [FREE Full text] [doi: 10.1155/2019/7681237] [Medline: 31093320]
6. Yang W, Wang H, Wu R, Lo C, Lin K. Home-based virtual reality balance training and conventional balance training in Parkinson's disease: a randomized controlled trial. J Formos Med Assoc 2016 Oct;115(9):734-743 [FREE Full text] [doi: 10.1016/j.jfma.2015.07.012] [Medline: 26279172]
7. Laver K, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
8. Cox DJ, Moore M, Burket R, Merkel RL, Mikami AY, Kovatchev B. Rebound effects with long-acting amphetamine or methylphenidate stimulant medication preparations among adolescent male drivers with attention-deficit/hyperactivity disorder. J Child Adolesc Psychopharmacol 2008 Mar;18(1):1-10. [doi: 10.1089/cap.2006.0141] [Medline: 18294083]
9. Pollak Y, Shomaly HB, Weiss PL, Rizzo AA, Gross-Tsur V. Methylphenidate effect in children with ADHD can be measured by an ecologically valid continuous performance test embedded in virtual reality. CNS Spectr 2010 Mar;15(2):125-130. [doi: 10.1017/s109285290002736x] [Medline: 20414157]
10. Barkley RA, Murphy KR, O'Connell T, Connor DF. Effects of two doses of methylphenidate on simulator driving performance in adults with attention deficit hyperactivity disorder. J Safety Res 2005;36(2):121-131. [doi: 10.1016/j.jsr.2005.01.001] [Medline: 15896352]
11. Montero-López E, Santos-Ruiz A, García-Ríos MC, Rodríguez-Blázquez R, Pérez-García M, Peralta-Ramírez MI. A virtual reality approach to the Trier Social Stress Test: Contrasting two distinct protocols. Behav Res Methods 2016 Mar;48(1):223-232. [doi: 10.3758/s13428-015-0565-4] [Medline: 25673321]
12. Diemer J, Mühlberger A, Pauli P, Zwanzger P. Virtual reality exposure in anxiety disorders: impact on psychophysiological reactivity. World J Biol Psychiatry 2014 Aug;15(6):427-442. [doi: 10.3109/15622975.2014.892632] [Medline: 24666248]
13. Kothgassner OD, Felnhofer A, Hlavacs H, Beutl L, Palme R, Kryspin-Exner I, et al. Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. Computers in Human Behavior 2016;62:124-135. [doi: 10.1016/j.chb.2016.03.081]
14. Waterlander WE, Jiang Y, Steenhuis IH, Mhurchu CN. Using a 3D virtual supermarket to measure food purchase behavior: a validation study. J Med Internet Res 2015 May 28;17(4):e107 [FREE Full text] [doi: 10.2196/jmir.3774] [Medline: 25921185]
15. Pericot-Valverde I, Germeroth LJ, Tiffany ST. The use of virtual reality in the production of cue-specific craving for cigarettes: a meta-analysis. Nicotine Tob Res 2016 May;18(5):538-546. [doi: 10.1093/ntr/ntv216] [Medline: 26453669]
16. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
17. Ung C, Menozzi M, Hartmann C, Siegrist M. Innovations in consumer research: the virtual food buffet. Food Qual Prefer 2018;63:12-17. [doi: 10.1016/j.foodqual.2017.07.007]
18. Allman-Farinelli M, Ijaz K, Tran H, Pallotta H, Ramos S, Liu J, et al. A virtual reality food court to study meal choices in youth: design and assessment of usability. JMIR Form Res 2019 Jan 9;3(1):e12456 [FREE Full text] [doi: 10.2196/12456] [Medline: 30684440]

19. Persky S, Goldring MR, Turner SA, Cohen RW, Kistler WD. Validity of assessing child feeding with virtual reality. Appetite 2018 Apr 1;123:201-207 [FREE Full text] [doi: 10.1016/j.appet.2017.12.007] [Medline: 29277518]
20. Moustafa F, Steed A. A Longitudinal Study of Small Group Interaction in Social Virtual Reality. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. 2018 Presented at: VRST'18; November 28-December 1, 2018; Tokyo, Japan. [doi: 10.1145/3281505.3281527]
21. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 2;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
22. Horing B, Newsome ND, Enck P, Babu SV, Muth ER. A virtual experimenter to increase standardization for the investigation of placebo effects. BMC Med Res Methodol 2016 Jul 18;16:84 [FREE Full text] [doi: 10.1186/s12874-016-0185-4] [Medline: 27430476]
23. Deladisma AM, Cohen M, Stevens A, Wagner P, Lok B, Bernard T, Association for Surgical Education. Do medical students respond empathetically to a virtual patient? Am J Surg 2007 Jul;193(6):756-760. [doi: 10.1016/j.amjsurg.2007.01.021] [Medline: 17512291]
24. Persky S, Ferrer RA, Klein WM. Nonverbal and paraverbal behavior in (simulated) medical visits related to genomics and weight: a role for emotion and race. J Behav Med 2016 Oct;39(5):804-814 [FREE Full text] [doi: 10.1007/s10865-016-9747-5] [Medline: 27146511]
25. Jimenez YA, Cumming S, Wang W, Stuart K, Thwaites DI, Lewis SJ. Patient education using virtual reality increases knowledge and positive experience for breast cancer patients undergoing radiation therapy. Support Care Cancer 2018 Aug;26(8):2879-2888. [doi: 10.1007/s00520-018-4114-4] [Medline: 29536200]
26. Pandrangi VC, Gaston B, Appelbaum NP, Albuquerque FC, Levy MM, Larson RA. The application of virtual reality in patient education. Ann Vasc Surg 2019 Aug;59:184-189. [doi: 10.1016/j.avsg.2019.01.015] [Medline: 31009725]
27. Gulhane A, Vyas A, Mitra R, Oruche R, Hoefer G, Valluripally S, et al. Security, Privacy and Safety Risk Assessment for Virtual Reality Learning Environment Applications. In: Proceedings of the 16th IEEE Annual Consumer Communications & Networking Conference. 2019 Presented at: CCNC'19; January 11-14, 2019; Las Vegas, NV.
28. Botella C, Serrano B, Baños RM, Garcia-Palacios A. Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: a review of its efficacy, the adequacy of the treatment protocol, and its acceptability. Neuropsychiatr Dis Treat 2015;11:2533-2545 [FREE Full text] [doi: 10.2147/NDT.S89542] [Medline: 26491332]
29. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
30. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]

Abbreviations

VCT: virtual clinical trials
VR: virtual reality

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by B Feodoroff, C Johnson; comments to author 01.10.19; revised version received 11.10.19; accepted 20.10.19; published 03.01.20.

Please cite as:
Persky S
A Virtual Home for the Virtual Clinical Trial
J Med Internet Res 2020;22(1):e15582
URL: https://www.jmir.org/2020/1/e15582
doi:10.2196/15582
PMID:31899455

Viewpoint

Big Data in Context: Addressing the Twin Perils of Data Absenteeism and Chauvinism in the Context of Health Disparities Research

Edmund W J Lee[1,2,3], BA, MA, PhD; Kasisomayajula Viswanath[1,2], PhD

[1]Dana-Farber Cancer Institute, Boston, MA, United States

[2]Harvard TH Chan School of Public Health, Boston, MA, United States

[3]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore

Corresponding Author:
Edmund W J Lee, BA, MA, PhD
Dana-Farber Cancer Institute
375 Longwood Avenue
Boston, MA, 02215
United States
Email: Edmund_Lee@dfci.harvard.edu

Abstract

Recent advances in the collection and processing of health data from multiple sources at scale—known as big data—have become appealing across public health domains. However, present discussions often do not thoroughly consider the implications of big data or health informatics in the context of continuing health disparities. The 2 key objectives of this paper were as follows: first, it introduced 2 main problems of health big data in the context of health disparities—data absenteeism (lack of representation from underprivileged groups) and data chauvinism (faith in the size of data without considerations for quality and contexts). Second, this paper suggested that health organizations should strive to go beyond the current fad and seek to understand and coordinate efforts across the surrounding societal-, organizational-, individual-, and data-level contexts in a realistic manner to leverage big data to address health disparities.

(J Med Internet Res 2020;22(1):e16377) doi:10.2196/16377

KEYWORDS

big data; artificial intelligence; health informatics; wearable electronic devices; mobile health; social media; electronic health records; digital divide; health disparities

Introduction

The emergence of big data platforms is showing promise in addressing many public health problems, such as predicting and managing the spread of global infectious diseases by drawing on real-time data on social media [1,2], empowering people to monitor their health through wearable technologies and interact with health care providers through patient portals [3,4]. Big data are defined as extensive datasets characterized by 5 *Vs*—v*olume* (size of the data), *velocity* (speed at which data are collected and processed), *variety* (types of data), *veracity* (trustworthiness of the data), and *value* (usefulness for decision making), which would require sophisticated computing infrastructure for storage, management, and analysis [5,6]. Although many are optimistic about big data in bringing significant improvements to individuals' health [7,8], others have argued that implementation of big data or health informatics interventions could increase health inequality [9].

This is because health organizations (eg, hospitals, nongovernmental organizations, federal public health agencies, and academic institutions) that want to incorporate the use of big data in their work with underprivileged groups may arguably face additional challenges beyond computational complexities. In addition to the day-to-day data wrangling and predictive optimization, organizations working in public health also need to deal with the challenges, such as difficulty in recruiting, retaining, and obtaining data from population groups that have suffered disproportionately from disease burden. It is thus important to consider what are the challenges that impede underprivileged groups in achieving equitable health outcomes. This is critical to the success of deploying big data solutions to improve the health of underprivileged groups, as they may not have the resources to access some of the communication

technologies (eg, wearable gadgets and smartphones) that are used as apparatus for health data collection, leading to a new digital data divide [10].

The objectives of this commentary were 2-fold. First, we have presented 2 emerging challenges—*data absenteeism* and *data chauvinism*—that could significantly dilute the effectiveness of big data and health informatics initiatives in health disparities research. Second, we have argued that organizations involved in public health work should strive to understand the collection and use of big data in different contexts and coordinate efforts across societal, organizational, individual, and data levels to effectively address health disparities.

The Perils of Big Data: Absenteeism and Chauvinism

Any attempts to draw on big data to address health disparities will face enormous challenges. Referring to underprivileged groups, a recent report by the United States Agency for International Development in identifying challenges to big data implementation in resource-poor settings underscored 2 main obstacles: the quantity and quality of data from the poor [11]. From the get-go, data from the poor are often not represented because of the lack of cyberinfrastructure in some locations or the poor might not have access to the technologies required for their data to be captured. Even if the poor are represented, the data are often messy and incomplete, and blind faith in these data points—even if they are voluminous—would lead to biased results and inaccurate interpretations [12]. These problems related to the quantity and quality of data are characterized as *data absenteeism* and *chauvinism*, respectively.

Data absenteeism describes an ironic phenomenon of data scarcity in a data-rich society, where data from underprivileged groups are not represented—or severely underrepresented—in the databases of health organizations [13]. For instance, a study on the diversity and representation of racial groups across 51 biobanks in the United States found that compared with the US census, there were statistically significantly lower enrollment numbers for Hispanics and Latinos (US census: 18%; selected biobanks: 7%), as well as Hawaiian and Pacific Islanders (US census: 0.2%; selected biobanks: 0.01%) [14]. In another example, a study using the Healthcare Effectiveness Data and Information Set (2012-2015) found that in the United States, less than half of Medicare, Medicaid, and commercial insurance plans reported complete or partially complete data on ethnicity of their members, although the Affordable Care Act specifically required population health surveys in federal health programs to collect and report items on race, ethnicity, and language as part of the drive to reduce health disparities [15]. In addition, research has shown that the users of health technologies (eg, mobile health apps) are more likely to be younger, be highly educated, and have higher levels of digital literacy skills [16], and it is not known how represented underprivileged groups are in health interventions using health technologies.

This is a major issue that has been repeatedly documented as underprivileged groups may be overlooked or may not have easy access to big data platforms or devices that are often used for collecting data [17], be it social media [18], smartphones [16], or internet patient portals [4]. Even if accessibility is not a problem, the underprivileged groups would still face additional barriers. For instance, a study on the use of internet patient portals by a population of diabetic adults in Northern California showed that racial minorities were more likely to request for password reset when accessing internet patient portals and logged on less, suggesting that even with access, they were still left behind [19].

Drawing from the ecological perspective of health, there is a myriad of societal, organizational, and individual factors that collectively explain why data from underprivileged groups are not represented [20]. On a societal level, *social determinants* (eg, education level and economic and employment status) and *communication inequalities*—unequal access to and use of communication technologies—are contributing factors to data absenteeism. For instance, underprivileged groups may struggle with having access to necessities and infrastructure, such as sanitation, water, and proper housing, and having the latest digital communication devices may not rank high compared with these basic necessities. In addition, many of these digital technologies (eg, wearable gadgets) used for collecting public health big data may simply be out of reach for groups from lower socioeconomic position because of the cost factor. From an organizational perspective, using big data to address health disparities may be perceived as a costly long-term investment, and many small health organizations at the community level that cater to underprivileged groups do not have the *capacity* (eg, comprehensive data architecture) or the human *capital* (eg, staff who know how best to turn data into insights to benefit the organization operationally) to do so. On the individual level, recent high-profile scandals on misuse of data on social media, such as the Facebook-Cambridge Analytica scandal [21], could further erode trust in health systems.

Data absenteeism has serious ramifications, and it has a profound impact on underprivileged groups even if the effects are not visible or tangible in the short run. As government and public health systems are increasingly using big data to automate solutions for decisions pertaining to who would get public assistance and financial aid, data absenteeism could further penalize the underprivileged groups. These groups that require the most financial assistance for health and medical services would not be in the very system to contribute to the development of the machine learning algorithms to identify them and further deprive them of the assistance they need.

The Perils of Data Chauvinism

The second peril that threatens big data's efficacy in addressing health disparities is the problem of data chauvinism. Data chauvinism is the overconfidence that the acquisition of (big) data alone would be the panacea to health disparities, without due consideration for ensuring data quality when collecting data from the underprivileged groups. Clearly, the weaknesses and cracks of data chauvinism are visible in the light of some of the failings of high-profile projects, such as the Google Flu Trend (GFT) study, which overestimated the prevalence of flu compared with the Centers for Disease Control and Prevention's

(CDC) estimates from traditional reports from laboratories [22]. Certainly, quantity is not synonymous with quality, as fundamental threats to validity and reliability, such as data noise, confounds, and spurious relationships, needed to be accounted for when designing big data research and solutions. In the context of health disparities, the efficacy of big data in building models to predict outcomes may come under threat because of biases, such as self-selection and the lack of generalizability, resulting in overfitting of data [23].

As machine learning algorithms are typically trained on a training set before being applied to the test set, if there are inherent biases—reflected in the absence or incomplete data—in the training set, it would severely compromise the quality and accuracy of the prediction outcomes. Such instances of quantified discrimination have real-world repercussions and may further punish the underprivileged groups. In Indiana's experiment with welfare eligibility automation, some from the underprivileged groups lost their Medicaid benefits because the algorithms wrongly diagnosed them as *failing to cooperate*, thus disqualifying them from receiving the benefits [24].

Toward Understanding Big Data in Context

Recognizing the twin perils of data absenteeism and data chauvinism in the context of big data use for health disparities research, what steps could organizations take to address them considering that many are moving toward the integration of big data solutions into their system? There are no obvious and easy solutions, but we suggest that health organizations should strive to go beyond solely cultivating computational competency and consider societal-, organizational-, individual, and data-level contexts when implementing big data research and solutions to address health disparities to avoid the pitfalls of data absenteeism and data chauvinism.

Societal-Level Context: Addressing Social Determinants and Communication Inequalities

First, to combat data absenteeism and data chauvinism when designing big data research or health informatics interventions, health organizations should seek to understand how societal-level contexts, such as social determinants and communication inequalities, are barriers to the underprivileged in reaping the benefits of big data. In the context of interventions using smartphones or wearable gadgets, researchers need to be mindful that providing access to digital devices does not fully remove structural obstacles for the underprivileged groups. Apart from the costs of purchasing digital devices, the poor would need to bear additional recurring costs that are often minute from the perspective of the average working class. These are known as *connection maintenance costs* [25], and they could be the time, energy, and money that the poor need to maintain the connection to digital devices. One example of such costs could be ensuring that bills are paid on time to ensure continuous internet or phone connectivity, which previous research has documented as the key impediment to successful adoption of electronic health (eHealth) interventions [26]. In addition, wearable gadgets and health apps often work best on the latest operating systems, and if the poor are not able to spend more money to get the latest gadgets to obtain the latest updates, they would be systematically left out. Without consideration for these costs, studies have shown that even with the provision of technology and internet access, the underprivileged groups still faced significant barriers in taking advantage of big data and new technologies that would significantly improve their health [27] if they are unable to pay for continuous access. Studies have documented that when the underprivileged groups were unable to pay their phone bills, it had severe ramifications, as frequent changes in phone numbers would result in disrupted care, leading to missed appointments and important paperwork (eg, insurance claims) deadlines [25].

To alleviate these latent costs, researchers should be mindful to factor in an additional budget to reduce the connection maintenance costs borne by the underprivileged groups, such as covering their cell phone bills for health app interventions. For instance, in a study examining health information seeking habits among the underprivileged groups, the researchers conducted a randomized controlled trial to examine if provision of home computers, broadband internet access, training in computer use, and a Web portal designed for low-literacy populations would significantly improve internet use [13]. The results showed that participants in the intervention group (ie, those who received computers, internet access, computer training, and a Web portal) were more likely to use the internet compared with the control group. This demonstrates that when researchers are mindful in addressing hidden costs (eg, bills for internet connection) that participants need to bear to be a part of big data research projects, it would significantly reduce structural barriers that prevent them from fully engaging with the research.

Organizational-Level Context: Forging Strategic Data Alliances

Next, one of the key strategies for health organizations—regardless if they are well resourced or not—is to take active steps to forge strategic data alliances with other organizations that leverage their comparative advantage and circumvent their own organizational constraints. For instance, although large health institutions may have the resources to implement big data solutions and research, they may not be as effective as community health centers in reaching out to the poor [28]. Small health organizations (eg, community health centers), on the contrary, may not have the necessary training or infrastructure to use big data. A recent study examining rural public health system leaders' data needs in Alaska, Idaho, Oregon, and Washington found that they were ill equipped in data management and had limited experience with data analysis [29]. However, they would be valuable to large health organizations because of their access, experience, expertise, and relationship of trust established with the underprivileged groups [30]. Although this is easier said than done, there are a few practical ways to do this:

- Create Communities of Practice (CoP) where health organizations could come together periodically (eg, annually) to share best practices of big data use in addressing health disparities, their challenges, and identify strategies to engage the underprivileged groups.
- Through the CoP, build a mentorship culture where personnel from organizations that are further along in their big data journey could mentor staff from health organizations that are getting started using big data for health disparities.

One potential example of a CoP is the recent launch of a US $100 million initiative by the Rockefeller Foundation and other global health partners that aim to specifically empower frontline community health workers with the most affordable and latest innovations in data science for improving health [31]. Part of the initiative would entail creating a knowledge and data sharing network where partnering countries could tap into a global team of data science experts committed to sharing of technical expertise and resources in the context of improving community health.

Individual-Level Context: Building Trust With the Underprivileged Groups

Recognizing that issues of privacy violation, loss of confidentiality, and data abuse [32] are some of the reasons at the individual level for mistrust and cynicism in how big data are used in the health care system, it is crucial that health organizations prioritize establishing trust with the underprivileged groups. To do so, health organizations should strengthen communication efforts such that literacy support should be provided for any informatics intervention [9], and the tangible benefits to participants and their communities should be made clear without jargon. Previous research that used eHealth interventions in community settings with people from underprivileged groups found that in-person presentations and personal contact with community members and organizations were the most effective in recruitment and participation [26]. In other words, the design of health big data research should incorporate *people-powered data collaboratives*, where end users or beneficiaries of health big data should be treated as stakeholders and brought to the table from the get-go to give them a stake in deciding how and when their data could be used on their own terms [33]. Eliciting a higher degree of participation and engagement from the underprivileged groups would strengthen relationships and cultivate a group identity and possibly a sense of belonging [34], thereby enhancing greater trust.

An example of this is the *All of Us* research program led by the National Institutes of Health in the United States, which aims to gather lifestyle, environmental, and biological data—Electronic Health Records (EHRs), blood samples, and information from wearables and surveys—from 1 million or more people from diverse groups in the United States to improve biomedical research to advance health [35]. To improve trust with participants, researchers provided participants access to their own data and the results of any laboratory tests they undertook [36]. In addition, the researchers sought participants' feedback (in addition to experts) when drafting guidelines and frameworks on how the data could be better communicated with others.

Data Context: Prioritize Science Over Data in the Use of Data Science

Although understanding societal-, organizational-, and individual-level contexts would address data absenteeism, what can researchers do to avoid falling into the trap of data chauvinism? Ultimately, researchers within health organizations should prioritize scientific rigor in their use of data. There are 3 practical ways to do so. First, researchers should balance the *a priori* rigor of scientific inquiry with a data-driven paradigm and understand the context in which one would perform better than the other. The a priori scientific inquiry is the traditional scientific hypothesis testing approach where researchers first develop a set of research questions and hypotheses and set out to mine data to verify their assumptions. The data-driven paradigm draws much from existing machine learning approaches that seek to mathematically detect patterns in the data through the process of data wrangling, as well as refining algorithms from training datasets so that it could effectively predict outcomes [23]. Although there is nothing inherently wrong with this data-centric method, the danger of the current big data hype is the move toward a puritanical pursuit of being data driven at the expense of crowding out subject or domain experts or common sense. In the case of GFT, perhaps by taking a step back and asking the fundamental question of how reliable search queries were in serving as leading indicators of realities, it might attenuate the way the Google engineers thought about designing the algorithms and thus avoid the serious inflation of results.

Second, part of emphasizing the rigor of science is to consider data from multiple sources. After all, big data are not only about the volume but also the variety of sources. In the GFT example, one of the pitfalls was implicit algorithmic snobbery, where data and algorithms from Google were treated as superior compared with lagged data from the CDC. If, in the first place, the Google algorithms were dynamically recalibrated with CDC data (despite their limitations), it could have avoided the problem of overestimation [22].

Finally, health organizations should take steps to implement a data quality assessment framework, where researchers could evaluate their data in the context of the big questions on health disparities they are addressing. In this data quality assessment framework, researchers should go beyond addressing questions on why or what variables have missing values and aim to answer how effective the data are in helping researchers address the root causes of health disparities. For instance, although the application of machine learning and artificial intelligence algorithms on EHRs may tell us which patients from underprivileged groups are more likely to get readmitted to hospitals for the same problem, the data would not empower health care providers to assess how best to alleviate the conditions to prevent readmissions. Thus, a rigorous quality assessment of big data in health disparities should guide researchers from simply asking, "what can these data tell us"

to "how can these data points reduce disparities and what additional data would be required?"

Conclusions

The potential to actualize the promises of big data in bridging health disparities to some extent is contingent on health organizations' efforts to address data absenteeism and data chauvinism. Although there are no easy solutions, it is crucial for health organizations to be keenly aware of both problems and develop a firm contextual understanding as well as coordinate strategies at the societal, organizational, individual, and data levels. Certainly, we agree that in the era of big data, taking small steps is crucial for success [37]; it also requires fundamental paradigm and attitudinal shifts within health organizations. Ironically, successful big data use in health disparities would require health organizations to look beyond data itself and to be intentionally inclusive so that no one is left behind so that the underprivileged could become the beneficiaries in the data revolution.

Conflicts of Interest

None declared.

References

1. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013;10(4):e1001413 [FREE Full text] [doi: 10.1371/journal.pmed.1001413] [Medline: 23565065]
2. Kuehn BM. Twitter streams fuel big data approaches to health forecasting. J Am Med Assoc 2015 Nov 17;314(19):2010-2012. [doi: 10.1001/jama.2015.12836] [Medline: 26575048]
3. Farrington C. Big data meets human health: Internet searches and fitness trackers are poised to play a role in the future of health care. Science 2016;353(6296):227 [FREE Full text] [doi: 10.1126/science.aaf8769]
4. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 1;24(5):927-932 [FREE Full text] [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
5. Wang Y, Kung L, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 2018;126:3-13. [doi: 10.1016/j.techfore.2015.12.019]
6. Chang WL, Grady N, NIST Big Data Public Working Group. NIST: National Institute of Standards and Technology.: National Institute of Standards and Technology; 2015. NIST Big Data Interoperability Framework: Volume 1, Big Data Definitions [Version 2] URL: https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-1-big-data-definitions-version-2 [accessed 2019-11-01]
7. Onukwugha E, Duru OK, Peprah E. Foreword: big data and its application in health disparities research. Ethn Dis 2017;27(2):69-72 [FREE Full text] [doi: 10.18865/ed.27.2.69] [Medline: 28439175]
8. Carney TJ, Kong AY. Leveraging health informatics to foster a smart systems response to health disparities and health equity challenges. J Biomed Inform 2017 Apr;68:184-189 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.011] [Medline: 28214562]
9. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
10. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010;15(Suppl 3):216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
11. Breen N, Jackson JS, Wood F, Wong DW, Zhang X. Translational health disparities research in a data-rich world. Am J Public Health 2019 Jan;109(S1):S41-S42. [doi: 10.2105/AJPH.2019.304951] [Medline: 30699034]
12. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. Practical guidance on artificial intelligence for health-care data. Lancet Digit Heal 2019;1(4):e157-e159. [doi: 10.1016/S2589-7500(19)30084-6]
13. 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]
14. Cohn EG, Hamilton N, Larson EL, Williams JK. Self-reported race and ethnicity of US biobank participants compared to the US Census. J Community Genet 2017 Jul;8(3):229-238 [FREE Full text] [doi: 10.1007/s12687-017-0308-6] [Medline: 28623623]
15. Ng JH, Ye F, Ward LM, Haffer SC, Scholle SH. Data on race, ethnicity, and language largely incomplete for managed care plan members. Health Aff (Millwood) 2017 Mar 01;36(3):548-552. [doi: 10.1377/hlthaff.2016.1044] [Medline: 28264958]
16. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
17. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):S131-S133 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
18. Hargittai E. Potential biases in big data: Omitted voices on social media. Soc Sci Comput Rev 2018:1-15. [doi: 10.1177/0894439318788322]

19. 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]
20. Moran MB, Frank LB, Zhao N, Gonzalez C, Thainiyom P, Murphy ST, et al. An argument for ecological research and intervention in health communication. J Health Commun 2016;21(2):135-138 [FREE Full text] [doi: 10.1080/10810730.2015.1128021] [Medline: 26788682]
21. Confessore N. The New York Times. 2018 Apr 4. Cambridge Analytica and Facebook: The Scandal and the fallout so far URL: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [accessed 2019-11-01]
22. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
23. Kreatsoulas C, Subramanian S. Machine learning in social epidemiology: learning from experience. SSM Popul Health 2018 Apr;4:347-349 [FREE Full text] [doi: 10.1016/j.ssmph.2018.03.007] [Medline: 29854919]
24. Featherson L. The New York Times. 2018 May 4. How Big Data is Automating Inequality URL: https://www.nytimes.com/2018/05/04/books/review/automating-inequality-virginia-eubanks.html [accessed 2019-11-01]
25. Gonzales AL, Ems L, Suri VR. Cell phone disconnection disrupts access to healthcare and health resources: a technology maintenance perspective. New Media Soc 2016;18(8):1422-1438. [doi: 10.1177/1461444814558670]
26. Nagler RH, Ramanadhan S, Minsky S, Viswanath K. Recruitment and retention for community-based eHealth interventions with populations of low socioeconomic position: strategies and challenges. J Commun 2013 Feb 1;63(1):201-220 [FREE Full text] [doi: 10.1111/jcom.12008] [Medline: 23439871]
27. 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 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
28. Boutin-Foster C, Scott E, Melendez J, Rodriguez A, Ramos R, Kanna B, et al. Ethical considerations for conducting health disparities research in community health centers: a social-ecological perspective. Am J Public Health 2013 Dec;103(12):2179-2184. [doi: 10.2105/AJPH.2013.301599] [Medline: 24134347]
29. Bekemeier B, Park S, Backonja U, Ornelas I, Turner AM. Data, capacity-building, and training needs to address rural health inequities in the Northwest United States: a qualitative study. J Am Med Inform Assoc 2019 Aug 1;26(8-9):825-834. [doi: 10.1093/jamia/ocz037] [Medline: 30990561]
30. Stephens KK, Rimal RN, Flora JA. Expanding the reach of health campaigns: community organizations as meta-channels for the dissemination of health information. J Health Commun 2004;9(Suppl 1):97-111. [doi: 10.1080/10810730490271557] [Medline: 14960406]
31. The Rockefeller Foundation. The Rockefeller Foundation. 2019 Sep 25. Using data to save lives: The Rockefeller Foundation and partners launch $100 million precision public health initiative URL: https://www.rockefellerfoundation.org/about-us/news-media/using-data-save-lives-rockefeller-foundation-partners-launch-100-million-precision-public-health-initiative/ [accessed 2019-11-01]
32. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis 2017;27(2):95-106 [FREE Full text] [doi: 10.18865/ed.27.2.95] [Medline: 28439179]
33. Evans B, Krumholz H. People-powered data collaboratives: fueling data science with the health-related experiences of individuals. J Am Med Inform Assoc 2019 Feb 1;26(2):159-161 [FREE Full text] [doi: 10.1093/jamia/ocy159] [Medline: 30576557]
34. Ramanadhan S, Mendez SR, Rao M, Viswanath K. Social media use by community-based organizations conducting health promotion: a content analysis. BMC Public Health 2013 Dec 5;13:1129 [FREE Full text] [doi: 10.1186/1471-2458-13-1129] [Medline: 24313999]
35. National Institutes of Health. 2019. All of Us Research Program Overview URL: https://allofus.nih.gov/about/all-us-research-program-overview [accessed 2019-11-01]
36. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The 'All of Us' Research Program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
37. Heller N, Seltzer JH. Commentary: in the rising era of big data, small steps are key. Ethn Dis 2017;27(2):73-76 [FREE Full text] [doi: 10.18865/ed.27.2.73] [Medline: 28439176]

Abbreviations

CDC: Centers for Disease Control and Prevention
CoP: Communities of Practice
eHealth: electronic health
EHR: electronic health record
GFT: Google Flu Trend

Edited by G Eysenbach; submitted 23.09.19; peer-reviewed by N Zhao, C Fincham, L Guan; comments to author 29.10.19; revised version received 11.11.19; accepted 11.11.19; published 07.01.20.

Please cite as:
Lee EWJ, Viswanath K
Big Data in Context: Addressing the Twin Perils of Data Absenteeism and Chauvinism in the Context of Health Disparities Research
J Med Internet Res 2020;22(1):e16377
URL: https://www.jmir.org/2020/1/e16377
doi:10.2196/16377
PMID:31909724

Viewpoint

Barriers to Working With National Health Service England's Open Data

Seb Bacon[1], BA; Ben Goldacre[1], BA, MA, MSc, MBBS, MRCPsych

The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Ben Goldacre, BA, MA, MSc, MBBS, MRCPsych
The DataLab
Nuffield Department of Primary Care Health Sciences
University of Oxford
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289313
Email: ben.goldacre@phc.ox.ac.uk

Abstract

Open data is information made freely available to third parties in structured formats without restrictive licensing conditions, permitting commercial and noncommercial organizations to innovate. In the context of National Health Service (NHS) data, this is intended to improve patient outcomes and efficiency. EBM DataLab is a research group with a focus on online tools which turn our research findings into actionable monthly outputs. We regularly import and process more than 15 different NHS open datasets to deliver OpenPrescribing.net, one of the most high-impact use cases for NHS England's open data, with over 15,000 unique users each month. In this paper, we have described the many breaches of best practices around NHS open data that we have encountered. Examples include datasets that repeatedly change location without warning or forwarding; datasets that are needlessly behind a "CAPTCHA" and so cannot be automatically downloaded; longitudinal datasets that change their structure without warning or documentation; near-duplicate datasets with unexplained differences; datasets that are impossible to locate, and thus may or may not exist; poor or absent documentation; and withholding of data for dubious reasons. We propose new open ways of working that will support better analytics for all users of the NHS. These include better curation, better documentation, and systems for better dialogue with technical teams.

(J Med Internet Res 2020;22(1):e15603) doi:10.2196/15603

KEYWORDS

informatics; health services; software; access to information

Background

Open data is briefly defined as data that anyone can access, use, modify, and share; more technical definitions are available from various sources [1]. Open data must, therefore, be shared in the public domain or provided under an open license, accessible and downloadable without charge, provided in a form that is machine readable, and provided in an open format, which itself places no restrictions on use.

The UK government has long recognized that simply publishing data is, in itself, not sufficient to meet these criteria and also not sufficient to drive change and innovation. The 2012 Open Data White Paper set out 14 information principles declaring that data should be easy to find, available without registration, and accompanied by meaningful descriptive text, alongside various other more technical recommendations [2]. It also adopted the 5-star scheme from Tim Berners-Lee for assessing the extent to which datasets are truly reusable: this ranges from unstructured proprietary documents through to fully linked data in nonproprietary formats with uniform resource identifiers. Although a swathe of data has been licensed appropriately for free reuse, the more detailed principles outlined in the White Paper have not been consistently adopted.

The National Health Service (NHS) in England has long agreed that transparency can lead to better outcomes for patients and taxpayers [3,4], and the Department of Health first requested publication of some prescriptions data as early as 1998 [5]. However, to our knowledge, there has been very little work describing how open health data are used in practice by NHS analysts, industry, or health researchers; and no prior work on the barriers was encountered.

Our group develops and maintains OpenPrescribing.net, an online and publicly accessible tool, to help users explore highly granular NHS primary care prescribing open data. It is widely used with over 15,000 unique users each month. Its users are predominantly from within the NHS, but industry and patient groups are also well represented. In England, the planning and commissioning of health care services for each local area is carried out by Clinical Commissioning Groups (CCGs), who alongside NHS England commission primary care services from individual general practitioner (GP) practices. GPs have considerable freedom in prescribing behavior, with the costs of prescriptions usually being borne by CCGs. The transfer of money from CCGs to pharmacies (and other organizations) who dispense prescriptions to patients is mediated by the NHS Business Services Authority (NHSBSA), which processes all prescribing transactions to determine correct payments. The NHSBSA is, therefore, also responsible for converting data submitted from pharmacies into a standard format. Although it exists for an economic purpose, the existence of this very high-quality dataset provides a unique opportunity to find ways to improve the quality, safety, and cost-effectiveness at all individual GP practices across England. Our tools support complex bespoke data queries alongside numerous predefined standard measures for safety, cost, and effectiveness. In total, 92.1% (176/191) of CCGs are signed up to monthly alerts, which automatically identify high-priority action items. We have published peer-reviewed research showing that prescribing is substantially improved in practices and CCGs where OpenPrescribing.net data are accessed [6].

OpenPrescribing.net is built on top of data that are theoretically publicly accessible. We have repeatedly encountered time-consuming barriers to accessing and processing these data. In this paper, we have described some of these barriers and made recommendations on how the NHS could share data more effectively.

The views set out below are informed by our technical work building OpenPrescribing.net but also by our broader background. The DataLab at the University of Oxford is a mixed team of software engineers, clinicians, academics, and analysts turning NHS data into tools and services to directly improve patient care. We aim to pool skills and combine best practices from software engineering and academia, producing open source software, open prototypes, and open workbooks. On GitHub, under open licenses, we have shared 44,000 lines of code in 34 public repositories with over 5000 commits; 850 Python files; 105 Structured Query Language (SQL) files containing 4600 lines of SQL; 140 Jupyter notebooks; and over 1000 GitHub issues, each containing detailed descriptions of specific problems we have encountered and their technical solutions. Many of us have also worked previously in organizations that promote open access to knowledge. In more concrete terms, as reference to our experience of working with NHS open data, at least 8 different datasets must be located, downloaded, converted, normalized, interpreted, combined, and then processed to create even 1, apparently simple, mapped insight on OpenPrescribing.net: “over the past 5 years, NHS North Cumbria spent £63,000 on Linaclotide.”

In the following section we have described a range of barriers we have encountered in accessing NHS open data. For each problem domain we describe the datasets we are aiming to access, the barriers encountered, and some suggested solutions that would make the data usable and impactful.

Problems With the Prescribing Data Itself

Each month, we download and process prescribing data for NHS England. The best practice [7-9] is that this should be easily discoverable, accessible without human intervention, made available at addresses that do not change, and documented so the relevant concepts are clearly explained. None of these are entirely true of the prescribing data.

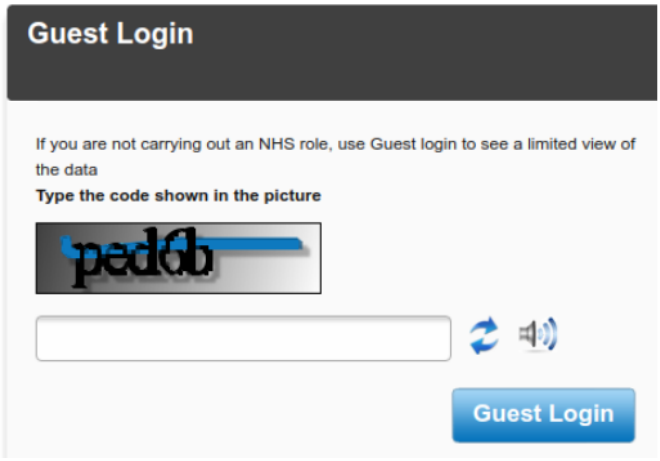
For example, consistently locating the data is difficult: both initially and on an ongoing basis with each new month of data. The first challenge for a consumer of the data is picking a dataset to use. A total of 2 very similar datasets are published by 2 different organizations: NHS Digital, and NHSBSA. The NHS Digital dataset is published on the first Friday of the third month after data collection, whereas the NHSBSA dataset is usually available 6 weeks following data collection. Neither of these datasets reference the other in their documentation, and we have found no single location that identifies them as complementary sources.

Until 2017, we used NHS Digital’s version (known as practice level prescribing data), simply because this is the easiest to find. For 2 years we retrieved the data from NHS Digital’s data repository [10] but that link broke during 2018; following a content reorganization, it is now available on a new NHS Digital website [11]. Complicating easy discovery of this dataset is the fact that it is also listed in NHS England’s *Data Catalogue* but only with data up to May 2016 [12]. Every time the location of the data changes, it breaks the software we have written to automatically download it.

The version that we have used since 2017 is known as Practice Detailed Prescribing Information (PDPI) and is published by NHSBSA on their Information Services portal [13]. The decision to switch to using that dataset was driven by its timeliness: the monthly release is available much sooner than the practice level prescribing data. However, it is difficult to use. First, the dataset is only accessible after completing a CAPTCHA, which means automating the download process is impossible: every month a software engineer has to manually fill out a form (Figure 1).

Second, although documentation is provided for PDPI, the documentation is incomplete: it refers to fields that do not exist, and does not refer to 15 fields that do exist [14]. Finally, to our surprise, we read in a newsletter from NHSBSA that the version of the data on their Information Services portal would be replaced by a new system in December 2018 [15]. We have since established that nothing will change for the time being for end users but were surprised there was no public consultation about the possibility. Changing our systems to support a new location (and potentially format) could conceivably take several weeks and early warning for this kind of change is essential. Textbox 1 outlines some steps that could be taken to improve access to data.

Figure 1. CAPTCHA form for National Health Service Business Services Authority Practice Detailed Prescribing Information dataset.



Textbox 1. How access to data could be improved.

- No publicly available data should be protected behind a CAPTCHA.
- Each dataset should have every field documented.
- Every resource should have a consistent location (URL or machine-readable data index) for finding current data.
- Internal reorganizations should not result in these URLs being deleted; if they are superseded, old URLs should be kept and set up to redirect to new locations.
- When there is a plan to relocate or change datasets, this should be advertised and documented well in advance.
- It should be easy to find all current prescribing data resources and to pick the most appropriate one. For example, there could be a single place listing all current prescribing data resources.

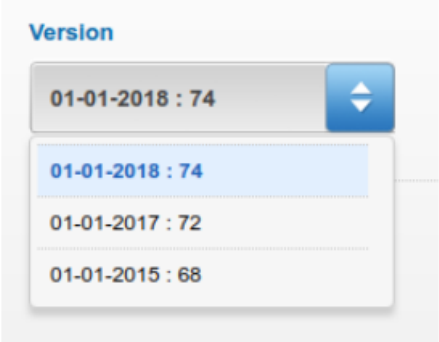
British National Formulary Names

Each prescription is identified by a "BNF Code": this is typically 15 characters long and uniquely identifies a presentation of a drug. For example, the code for Tramadol HCl 300 mg tablets is 040702040AAAMAM. To make prescribing data useful for analysts, all the British National Formulary (BNF) codes must be converted to human-readable BNF names. Data to support this are published by NHSBSA (behind another CAPTCHA) on the Information Services portal [16].

The coding scheme is based on the BNF's old classification, which they no longer maintain themselves. Therefore, the NHS' altered version is properly known as the "Pseudo BNF Classification" [17]. Changes to BNF coding take place throughout the year, with a large reclassification process happening every January, when some drugs are moved between BNF sections or BNF chapters, and others are given entirely new BNF codes. This reclassification process is not mentioned, let alone described, anywhere we can find on the internet. The process by which the BNF file is updated is unclear. Although we know a major revision is published every January, minor revisions are also published monthly, but a user would not know this because the data download page only refers to the January editions (Figure 2).

The fact that some BNF codes change over time makes time-based analysis of data difficult. For example, a user searching for Linaclotide, using its current BNF code, will find no prescribing before 2014. This is because the drug was moved from BNF section *1.2: Antispasmod. & Other Drgs Alt. Gut Motility* to BNF paragraph *1.6.7: Other Drugs Used In Constipation*, and its BNF code changed accordingly.

Figure 2. British National Formulary Code data labeled available in November 2018.



As there is no mention that such changes are possible on the internet, we first inferred this was the case following user enquiries about apparently disappearing drugs. Following direct enquiries, we now obtain a spreadsheet detailing the changes every January by emailing NHSBSA directly and apply this to the imported prescribing data. By comparing the pseudo BNF code lists each month, we have inferred that codes also sometimes change mid-year but have not yet obtained access to these individual changes on a monthly basis [18]. Textbox 2 describes some of the ways the NHS could aid public understanding of BNF code changes.

Textbox 2. How British National Formulary change management could be improved.

- Published, open data should never be protected behind a CAPTCHA.
- The fact that the British National Formulary (BNF) scheme changes regularly should be documented.
- There should be a single, obvious channel for data consumers to query possible issues in the data.
- BNF code changes should be published monthly as a mapping.
- Each data release should be clearly labeled on its index page, so users know when a new version has been released; there should be a way for users to subscribe to be notified of new releases.

Practice Data

To include patient list size in our analyses, and show practice names and addresses, we looked up extra information in a dataset provided by NHS Digital. The format of this dataset has not changed since 2015; however, we have encountered regular problems with its location changing, which has prevented us from fully automating this monthly process.

Until 2018, our procedure was to automate a search for the phrase "*Number of Patients Registered at a GP Practice*" on the NHS Digital website and then look for datasets in the list of results returned. From July onward, the data were moved to a different location. In addition, the format of the dates within the file changed between June and July. The location has changed twice more since then. All these changes mean that it is common for the code that automatically imports practice list information to break and to require manual input.

Once practice data were obtained, we encountered difficulties with data quality. In general, the data provided by NHSBSA are of a high standard. However, there is no documentation for several known recurring errors and no way to report and correct them systematically.

For example, it is important to know whether an institution is a standard GP practice or a different kind of institution (eg, a homeless service or a drop-in center). However, in the data provided, there is a small but significant number of obvious errors in coding, such as classification of care homes [19] and violent patient services [20] as standard settings. When we queried these problems, we were informed that errors can only be corrected by CCGs themselves; however, they were unable to provide us with CCG contact details to contact these organizations ourselves systematically and notify them of the need to make these corrections to their own data. It is also unclear if there is any part of the NHS that considers itself responsible for maintaining accurate data in this area.

As a final example, this problem is further compounded by list size data that regularly appears to contain fictional values. Sometimes we identify practices that have prescribing at improbable levels, far exceeding their total number of patients [21]. These may be data entry errors but sometimes appear to be caused by an unusual design of the data specification: when a new practice is registered, the NHSBSA proforma states that a list size *must* be given, which can be "nominal" but must be under 100 [22]. Our interpretation of this is that any list size of less than 100 must be considered arbitrary and cannot be relied on. This interpretation may be wrong but is our best guess in the absence of documentation. The best practice for data management is that missing values should be clearly coded as such. Textbox 3 contains further suggestions for aiding consumption of practice list size data.

Textbox 3. How practice data quality and accessibility could be improved.

- All data should be published in a predictable format and location.
- "Nominal" values should not be used: missing values should be clearly coded as "missing."
- Where there are systemic issues with data quality, these should be documented.
- There should be a clearly documented and centralized system for reporting and correcting errors in the data.
- Data stewards should take responsibility for collecting error reports and aim to correct them.

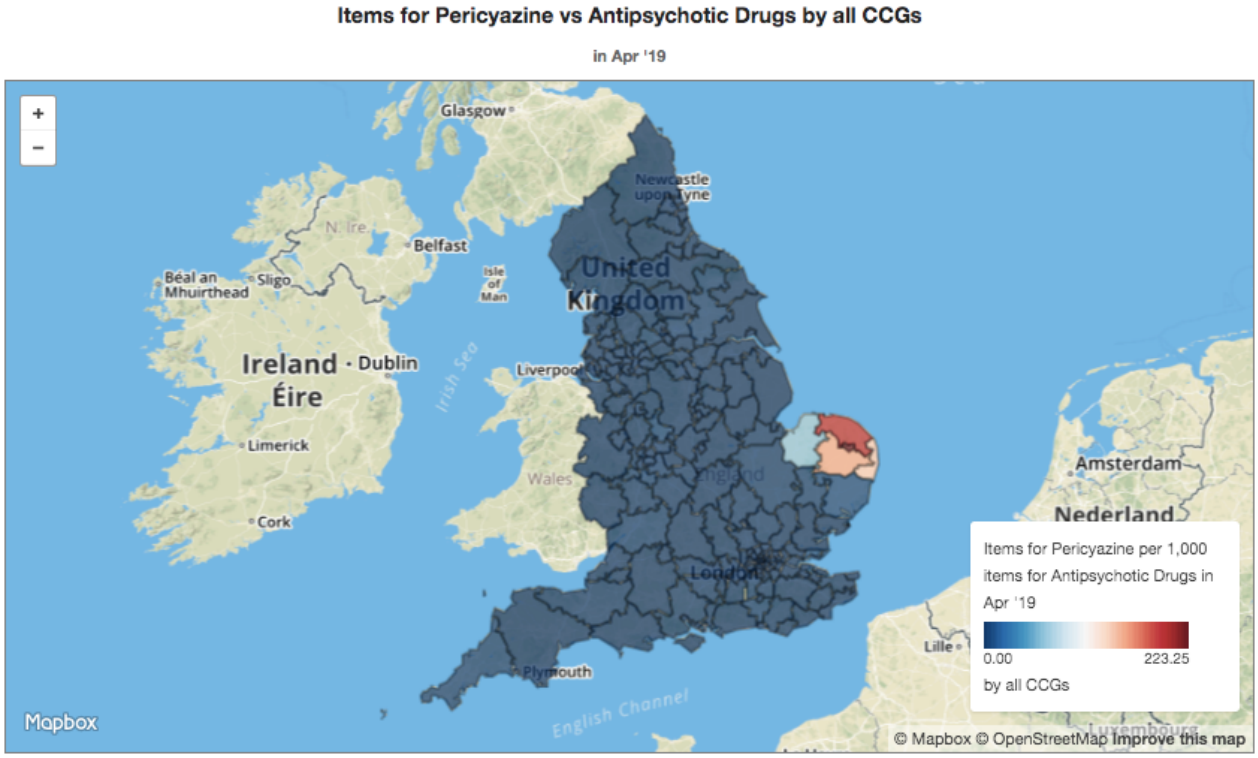
Clinical Commissioning Group Codes, Boundaries, and Membership

To analyze data at a CCG level, we need to aggregate the per-practice data up to CCG level. The source data provide a CCG for each row, so this is straightforward for contemporary data.

Maps

We show CCG boundaries on maps in various places in OpenPrescribing.net: an example in Figure 3 shows unusually high prescribing of pericyazine, a very unusual antipsychotic, in Norfolk. Until 2017, we obtained the map data from NHS England [23]. This was updated occasionally; on one occasion it was supplied in a different format from usual, so we had to alter our software accordingly. In 2017 it became apparent that this map was no longer being kept up to date with changes in CCGs. We found a new file provided by the Office for National Statistics (ONS) [24], which was up to date. For some time, both maps were available. During this period, the NHS England map was 2 years out of date, and with no indication there was a more up-to-date map in a different location. Eventually, the NHS England version disappeared from the website. The current ONS map appears to be accurate but, unlike the NHS England one, does not include a CCG identifier, so a separate data file is needed to associate CCGs with their boundaries; again, the file is supplied with no clear indication of where such a mapping can be found.

Figure 3. Prescribing of pericyazine as a proportion of all antipsychotics across all Clinical Commissioning Groups in England, as displayed on OpenPrescribing.net.



Clinical Commissioning Group Practice Membership Changes

Our per-practice data provides a practice's CCG membership for the current month. However, historic analysis is complicated by the fact that practices often change CCGs, CCG boundaries sometimes change, and CCGs often merge. In 2017, for example, the boundary between NHS Cumbria and NHS Morecambe Bay changed. We were able to infer from the data that 32 practices moved to Morecambe Bay as a result [25].

The problem that a practice may move between CCGs is addressed in the OpenPrescribing.net software by projecting the practice's current CCG membership back in time: for example, a prescription dispensed in 2012 is allocated to whatever CCG that practice currently belongs to. This works well in most cases but becomes complicated when a practice has closed. In the case of Cumbria in April 2017, 5 practices had closed before the boundary change; these are, therefore, not reflected in current CCG membership data. This leaves the problem of which CCG to attribute them to: their patients have not disappeared, just moved, but it is impossible to find out or infer where they were moved to because the information about what happens to a practice's patient list on closure is not available as data.

Our own research has established that when a practice closes (or merges), it must fill out at least two nearly identical forms to notify the prescription pricing division of NHSBSA [26] and Primary Care Support England (PCSE) [27]. We requested any data resulting from both these forms in Freedom of Information (FOI) requests to the respective organizations [28,29]. NHSBSA informed us that "Prescriber and practice/cost centre amendments are only held as paper forms," and PCSE eventually supplied us with a spreadsheet that appeared to bear little relationship to real closures [30]. As a result, every April (when boundary changes happen), our developers have to make educated guesses about which contemporary CCGs the patients of a closed practice now belong to, and at a practice level, there is nothing we can do to amend the data correctly [25].

Textbox 4 provides some suggestions on how the quality and accessibility of mapping data could be improved.

Textbox 4. How National Health Service mapping data could be improved.

- Map files should be published in a single, easily found, permanent location to a regular schedule.
- They should be published alongside (or indicate the location of) files supporting mapping to standard National Health Service (NHS) clinical commissioning group (CCG) codes as used in prescribing data.
- Their format should stay constant over time where possible.
- Practice merger and closure data should be published, showing where and when practice lists have transferred.
- Even if this is not possible, the problem of tracking historic prescribing behavior via practice codes should be clearly documented.
- We are unclear as to the value for the NHS of a system that requires CCGs to notify NHS England of practice changes but then leaves the information on paper.

What Does "Quantity" Mean?

A single row of prescribing data includes a column denoting the quantity of the item dispensed. For example, in the case of paracetamol tablets, a "quantity" of 25 means that 25 *tablets* were dispensed. This field is used in most of our analyses. For example, our price-per-unit tool [6] identifies possible savings by comparing the price of dispensed drugs between practices nationally, at a "quantity" level. However, a consistent and precise definition of what "quantity" means for each product has been elusive [31]. For example, the NHS Digital glossary [17] defines quantity as follows:

> *The quantity of a drug dispensed is measured in units depending on the formulation of the product, which is given in the drug name. Where the formulation is tablet, capsule, ampoule, vial etc. the quantity will be the number of tablets, capsules, ampoules, vials etc. Where the formulation is a liquid, the quantity will be the number of millilitres. Where the formulation is a solid form (eg. Cream, gel, ointment), the quantity will be the number of grammes.*

However, this definition is not sufficiently precise for use in statistical analyses. For example, it is not obvious if a foam should be classified as a liquid or a solid. Further extensive investigation uncovered the existence of a "standard quantity unit" field for every product, which defines the property precisely. However, it can be found only in one place, the monthly prescription cost analysis spreadsheet [32] and is not mentioned anywhere outside that dataset. This useful column was removed without warning in the data from December 2018 onward.

Even when the standard quantity unit for a presentation is known, the definition of quantity sometimes varies, for example, between "dose" and "pack." During development of our price-per-unit tool, we found a number of products where the highest price was orders of magnitude outside the normal range [33]. Items dispensed in packs of 56, for example, were sometimes being recorded as a quantity of both 1 or 56.

An NHSBSA glossary has this to say on the matter: "Where a product is packed in a 'special container'...*in some circumstances* [our emphasis] these items show quantity as the number of units supplied ie 1 or 2 even though a pack may contain 56 tablets" [31].

It is not clear from this statement if variation in the meaning of "quantity" for a single presentation is intentional, or accidental. We raised specific examples with NHSBSA, and this led to some of these figures being corrected retrospectively, but in other cases, we were told "work is under way to review this and agree a way forward."

Errors in data are inevitable and to be expected. Overall, the NHSBSA dataset is remarkably free of errors. However, as analysts we need to understand where errors are and, if they are systematic, where, and how often we can expect them to appear. The detailed investigative analysis required to understand these data delayed the launch of our price-per-unit savings tool by several months. This kind of delay has real-world effects; published peer-reviewed data show that the tool saves CCGs millions of pounds a year [6]. Textbox 5 summarizes some ways the meaning and quality of published datasets could be improved.

Textbox 5. How the meaning and quality of datasets could be made clear.

- By default, all prescribing data used internally at National Health Service Business Services Authority should be made available and described in 1 place.
- All data should be accompanied by clear, user-focused documentation about the meaning of each field.
- Where there are known problems with the data, these should be documented clearly and transparently.

How Can We Contact Practices by Email?

During 2017, we set out to conduct a simple, low-cost randomized trial: we notified GPs of cost-saving and quality improvement opportunities in their prescribing and are currently measuring the impact of this notification on their behavior. The intervention was split between 3 methods of communication: letter, email, and fax. We assumed there must be at least one central NHS database of practices' email addresses; for example, NHS England emails a monthly GP practice bulletin to GP practices. We knew there might be problems making this public,

but we were also surprised by how difficult it was to find out if the database existed at all.

First, we checked WhatDoTheyKnow, a publicly accessible archive of requests made under the FOI Act for any past FOI requests for GP practice contact information. We found NHS England had refused a similar request for practice information stating that "NHS England does not hold information in relation to your request" [34]. We knew this was unlikely to be correct, so we sent a new request asking specifically for the contact details for the GP practice bulletin, which we knew was emailed to practices by NHS England [35]. The response acknowledged the existence of a list and recognized there is "a general public interest in the release of such information in-line with NHS England's commitment to openness and transparency."

However, the request was refused under 2 of the allowed exemptions in the FOI Act. The first was section 40 (an exemption relating to personal information). They argued it would be unfair to staff, who had signed up for one purpose, to be contacted for another purpose. The second was section 43 (an exemption relating to commercial interests). This is apparently because some of the GP email addresses had been purchased by NHS England from a third party under a license that forbids the NHS to share the information.

Having failed with one database we knew to exist, we made requests to every public body that might hold a database of GP email addresses. We preemptively included an argument that section 40 should not apply as these are work email addresses. All were refused, with similar arguments to those from NHS England or invoking section 21 (the information was already available—which is incorrect) or stating that they did not hold the information. The responses are summarized in Table 1.

Table 1. Summary of responses to Freedom of Information requests for general practitioner's email addresses.

Body	Reasons for not supplying the data
Department of Health and Social Care	Information not held [36]
NHS[a] England (new request)	S21 [34]
NHS England (follow-up)	S40 [34]
Medicines and Healthcare Products Regulatory Agency	S21, S40, S43 (their own commercial interests) [37]
NHS Business Services Authority	S21, S40, S43 [38]
NHS Digital	S21 and information not held [39]

[a]NHS: National Health Service.

In our view many of the responses gave the impression of an organization actively seeking ways to refuse releasing this information. NHSBSA argued that providing email addresses would damage commercial interests because it decreases security: "The e-mail addresses could be used by cyber criminals to target practices, CCGs etc. If such an attack was successful it could result in financial loss and/or loss of patient data." This strikes us as an extremely unrealistic concern. The notion that hiding information intrinsically increases security has been long debunked in the security research community, where it is known disparagingly as "security through obscurity;" and in any case, the email addresses are all available through commercial data providers. Furthermore, most GP practices would expect to be contactable through email by their patients.

We were eventually able to run the randomized controlled trial (RCT) but only at greatly increased cost. We sent FOI requests to all 201 CCGs [40], of which 29 agreed to share at least some emails. We also wrote code to download data from the NHS Choices website. Finally, we combined our results with the commercial dataset that we purchased. In the end, we got emails for 27.0% (190/703) of practices from the NHS Choices website, 7.9% (56/703) from our FOI requests, and the remainder from a commercial provider. Notably, the email addresses purchased from a commercial vendor were apparently higher quality than those available directly from NHS resources: where we obtained an email address commercially, 18% (86/474) of practices accessed an email link versus 11% (23/211) for email addresses we obtained from other sources.

In summary, the reasons given for not supplying email addresses were inconsistent and sometimes hard to fathom. We believe the section 40 exemption (that it is unfair to individuals to release this data) is overused: given these are work addresses for managers, there would be a strong case for their release, based on current FOI guidelines. At the very least, given that 33% (66/201) of CCGs were willing to provide email addresses, the exemption is very unevenly interpreted and applied. We also understand that the vast majority of practices have a generic nhs.net inbox, which would certainly be exempt from section 40, but a list of even these email addresses is apparently unavailable.

Ultimately, it should not be difficult for researchers, health professionals, or indeed the public to have a way to contact any GP practice by email; and until this is the case, it should not be difficult (as it currently is) for a data consumer to establish definitively that there is no such resource. The problems we had assembling these data delayed the start of our RCT by several months. This delay indirectly affects care, as there is currently limited research available about how information is best disseminated through the NHS. We also note that the Secretary of State for Health and Social Care has prominently promoted the principle of using emails first, rather than letters, to communicate in the NHS. This is made harder if NHS organizations themselves are failing to make email addresses easily available, or actively blocking access. Textbox 6 summarizes the steps we suggest could be taken to address this problem.

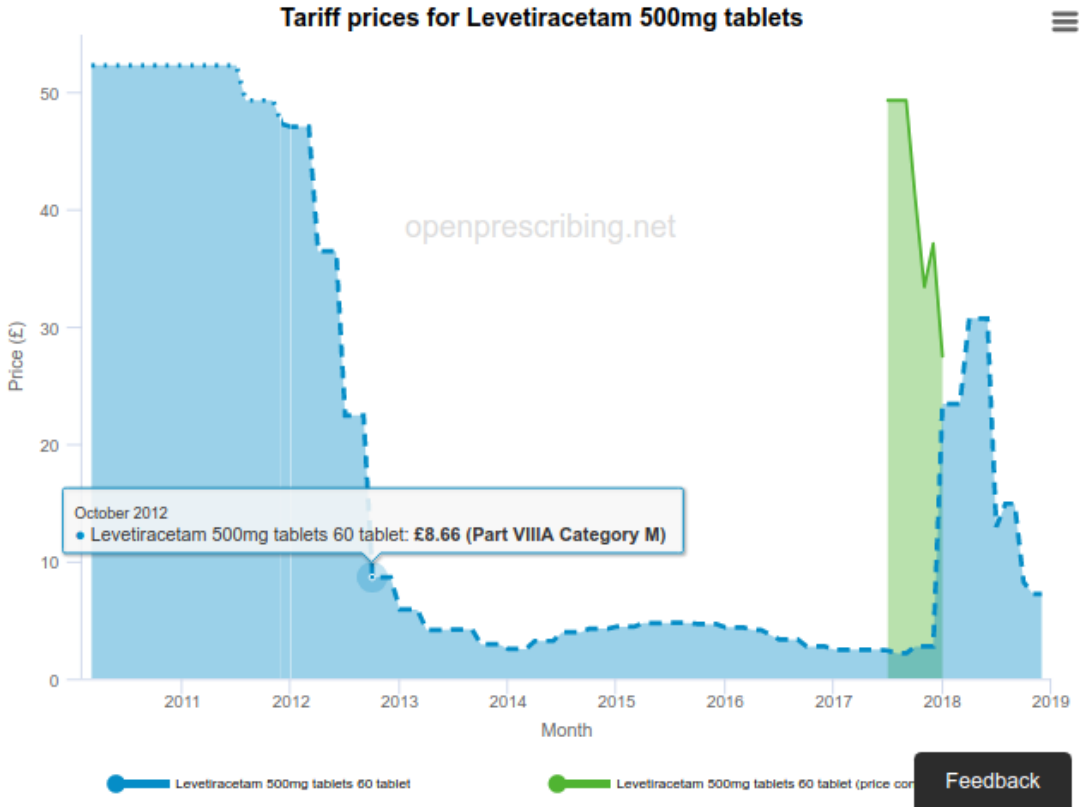
Textbox 6. How to make it easier to contact general practitioners.

- There should be a contact database for general practitioner practice managers, including email addresses, which is available to the public. Currently the choice to make an email address public is taken by practices alone.
- In the meantime, to save time and effort on the part of users of the data and NHS bodies, the fact that it is currently unavailable should be clearly documented in a single place, with an explanation, and suggestions for alternative sources.

What Are Historic Drug Tariff Prices?

For prescriptions written in primary care in England, the NHS reimburses community pharmacies for the medicines they purchase. The reimbursement price for generic medicines is set monthly by the Department of Health and Social Care (DHSC) in the NHS *Drug Tariff*. Price changes in the Drug Tariff are a major source of variation in costs for CCGs. In addition, the DHSC grants temporary excess rates in the form of price concessions every month, which can have sudden and unexpected effects on costs. We set out to create a tool on OpenPrescribing.net for tracking these changes over time and modelling the cost impact of new price concessions for each CCG using their past prescribing behavior (Figure 4). By processing the data as soon as it is published, we can provide email alerts to CCG budget holders warning them of upcoming price pressures.

Figure 4. Tariff prices and projected cost impact of price changes for Levetiracetam as on OpenPrescribing.net.



The tool was relatively easy to build. However, we could only build it after a large amount of difficult research and manual data editing. First, the data must be combined from spreadsheets found on 2 totally different websites, although both ultimately originate from DHSC. Second, each spreadsheet refers to the information in a different way, and they both provide separate files of data each month whose formats change over time. Finally, they are archived inconsistently, which makes it hard to locate historic data.

Finding the most recent data is relatively easy. Drug Tariff data are provided by the NHSBSA in a single location, which provides monthly spreadsheets for the last 2 years of the Drug Tariff [41]. In these data, every product is supplied with a unique Systematized Nomenclature of Medicine (SNOMED) code, which we can use to map to our prescribing data. Price concession data are published by the Pharmaceutical Services Negotiating Committee [42]; however, these data are supplied without SNOMED codes, which means the products must be matched to prescribing data by name only. This is hard to automate because the names often deviate in subtle ways (eg, “sq cm” vs “square cm”).

To find earlier datasets for previous years, we turned to the NHSBSA FOI archive [43]. The Web page has no search function, so we had to review all the content in the archive manually. We found 2 ZIP files of Excel spreadsheets in 2 different FOI requests [44,45]. These had gaps in their coverage. With educated guesses around filename and location, we were able to fill in these gaps by finding “hidden” files on the internet:

files that were available on the public internet and on the BSA website but which were not indexed in any search engines and were not linked to from anywhere. Having assembled the data, we had 49 files, in several different formats, which we standardized and combined using formulas, heuristics, and Excel plugins [46]. Once this was done, we were able to build a tool and automate monthly updates; however, because of the lack of SNOMED codes in the price concessions data, for most months we still had to manually match 1 or 2 products to the prescribing data. Textbox 7 summarizes some of the ways by which access to this (and similar linked datasets) could be improved.

Textbox 7. How price concessions data could be made easier to use.

- When archived data that are part of an already published time series are made available, the data should be published alongside the time series, not left in Freedom of Information requests.
- Any data export process of relatively small datasets should ideally involve producing a single file of all the data each month. This avoids the problem of formats changing through time.
- All data should be provided with identifiers.
- All shared data should be indexed or indexable.
- It is surprising that the price concession data are not already combined with Drug Tariff data somewhere in the National Health Service.
- Price concessions should be mentioned in documentation wherever Drug Tariffs are mentioned.

Discussion

Summary

As illustrated, although several NHS datasets are "open" by the narrowest definition, the NHS commonly breaches the principles of the Open Government White Paper and barely meets other best practice criteria such as the Berners-Lee 5-Star scale for open data. Collectively, the barriers described in this paper represent a substantial block to the development of new data-driven tools aiming to improve the quality, safety, and efficiency of NHS care. These barriers can be broadly divided into 4 areas: problems accessing the data, problems understanding the data, problems processing the data, and problems communicating with the NHS about the data. These imply 4 solutions: better curation, better documentation, better change management, and better dialogue with users. Here we summarize the barriers and offer some concrete suggestions of how the situation could be improved.

Better Curation

As documented in previous sections, there is a very substantial problem with curation of information in the NHS. Datasets are collected and shared at considerable expense but are then commonly undiscoverable or poorly indexed; they move location unpredictably, and often an interested user cannot establish whether a given dataset exists at all. The NHS England Data Catalogue [47] is largely unstructured: this means users must already know what they are looking for before they can find it. It also contains numerous older datasets, with no way for the user to deduce whether the dataset itself has been abandoned, with no further updates, or if only the catalog record is out of date. Catalogs are commonly divided by organization: this assumes that all users understand the complex organizational structure of the NHS and are able to predict whether a dataset is owned by NHS England, NHS Digital, NHS Improvement, NHSBSA, or some other organization. The problems we describe are often caused by ineffective automation. We propose 2 approaches for the NHS: "proactive curation" and "reactive curation."

Proactively, the NHS should invest in manually curating the data it already shares. This would entail detailed strategic input from experts in information management and librarianship; here we offer some brief principles. First, this curation should be done by people, with individuals or teams owning a particular topic area. Second, these teams must include domain experts already working within the relevant NHS organizations who understand the data. Third, instead of separate silos of data arranged by NHS organization, there should be a single location with links to all NHS data; and there should be confidence that all the data relevant to the topic are indexed in that one place as per best practice and government guidance [7]. Finally, these resources should be tagged in multiple dimensions, including their clinical domain, topic, and technical characteristics. In our view, the ideal model would be "topic-based guides" that are clearly branded, owned, and maintained by a single team; focused on adding new resources as they become available; and ensuring that data resources do not disappear or move.

Reactive curation also offers substantial benefits but at much lower cost. In short, where users are actively working on datasets, and they report to the NHS that something is missing, out of date, or poorly documented, then these errors, omissions, and shortcomings should be addressed and corrected. In short, there should be a simple means for users to report errors in the existing catalogs and for these errors to be corrected.

We can see no reason why any NHS resource should be behind a CAPTCHA, but if this is unavoidable, the reasons for this choice should be robustly documented and forewarning given in the catalog.

Better Documentation

In the previous sections we have documented numerous cases where NHS datasets are hard to interpret because of poor documentation and where the NHS has not been reactive to questions around poor or absent documentation. Documentation is challenging and time-consuming. However, good documentation brings clear thinking: an organization that cannot produce or share documentation on the data it holds is unlikely to be working effectively with that data internally.

We think there is room for the "proactive and reactive" model described above. Proactively, every dataset should be accompanied by documentation that explains its context (how and why it is used in the NHS), its provenance, the meaning of each field, how often it is updated, and any known issues with the data. At minimum, datasets that are regularly downloaded, used, or enquired about should be prioritized for this best practice. Reactively, the NHS should respond to queries, and there should be easy routes for users to give feedback on ambiguities or errors in the documentation. However, the NHS should also work more strategically with end users of the data, as this is where the true value of that dataset is often realized: documentation should ideally be developed in the open, in collaboration with data consumers, to ensure it is current and relevant.

Better Management of Change Over Time

As documented in previous sections, there are substantial problems with NHS datasets changing structure and format over time, often without those changes being documented. Often these changes are trivial: the names of the columns in a 2-way table or their order. However, every time the format of a dataset changes then there is a material consequence, for every end user: the pipelines for importing and processing data will break, the fault must be discovered, and developers must work around it. Commonly, there seems to be no technical reason for the changes we have seen in NHS datasets: it is likely that these changes simply reflect carelessness, or a lack of interest and knowledge about how the data are being used and processed by end users.

In an ideal world, data formats would never change; however, occasional changes are inevitable. Therefore, clear communication of changes is vital. We suggest that every dataset should be accompanied by a change log in its catalog entry. This change log should describe the nature of each format change. Crucially, there should also be clear documentation of the reasons why the change has happened, as this is likely to act as an informal feedback system, prompting NHS staff to think through whether the change is really necessary. There should be a way for consumers of the data to subscribe to updates and receive advance notice of these changes and to provide feedback where changes have happened without documentation, prompting the change log to be updated.

A related issue is stability of data structures over time when working with older datasets. Users often want to automatically retrieve and process not only current data but historic data. In doing so, they hit 2 problems: finding all historic files and resolving the format differences between them. To aid discovery, the naming conventions for data (eg, "title, date") should be documented in the same way as the data structure itself, and remain stable over time, to support automated retrieval. Where the formats of shared datasets must change, but the NHS holds historic data internally in a consistent format, then for all but the largest datasets, we suggest a bulk export of all historic data in the most current format should ideally be provided. Again, following the principle of transparency, where bulk exports are not possible, this should be mentioned and explained in the documentation.

Better Dialogue Between Data Producers and Data Consumers

We have returned repeatedly to the importance of better communication between data producers and consumers, if only as a means to reactively prioritize work around curation and documentation. In our view, this 2-way communication is vital: producers should be able to notify consumers of important changes in the data, and consumers should be able to notify producers of bugs in the data or ask questions. It is important to note here that good dialogue cannot be driven by a positive attitude alone: we have had many very positive interactions with single individuals in various NHS organizations who have been very helpful, but individuals can change jobs, or go on holiday, and finding the right person to talk to often relies on personal networks or sheer determination.

A strategic approach to dialogue requires good systems and formal structures. In short, the NHS needs a single place for users to ask questions about data, with the answers recorded and searchable in the public domain. This should be well publicized, open, public, and linked to liberally from across the NHS online estate. Given the NHS' general commitment to transparency [3,4], it would make sense for data producers in the NHS to borrow from best practices developed in the open source software movement, where source code is available to everyone, anyone can suggest edits, and anyone can report bugs. Each question answered in public will then be added to the commons of knowledge that can easily be accessed via any internet search engine. This platform should be curated by an NHS employee who has the authority to pursue questions on users' behalf and expect answers within a reasonable time frame. Over time, a database of questions and answers could evolve into a series of topic-based guides, written in collaboration with the user community. Many of the problems faced by those working with NHS data will already have been solved several times over by analysts within the NHS or third parties elsewhere.

None of this requires custom software and could all be provided through standard, widely used, free, open services such as GitHub and GitLab. End users could contribute bug reports to the documentation; they could ask questions through the bug tracking systems; everyone could see everyone else's input, helping raise standards and awareness; and the data producers could reply on built-in notification features to push feeds of updates to the end users. The most important part of solving this problem is not software but staff expertise and time. By reducing the friction between the 2 sides of the data exchange, better uses of data will emerge.

For clarity, this is not a "blue skies" or challenging suggestion. This is a standard way of working outside of the NHS, and it is how our own team works: we document every step of our problem solving publicly, in our closed and open "issues" on GitHub, which now number over 1000 [48]. Anyone working with NHS data who has been blocked by the same technical barriers we have hit can find our solutions—and the reasoning leading up to them—simply by using a search engine.

Conclusions

Releasing data under open licenses was the starting point for open data and the open government movement, in which the United Kingdom has been a global leader. However, in our experience, the implementation of these open principles in the NHS has been absent or flawed, with poor documentation, poor curation, and poor dialogue presenting substantial barriers to innovation. We have chosen to spend time documenting these issues at length; many third parties confronted with similar barriers will either give up, concluding a service as impractical, or quietly expend substantial resource and effort on workarounds. This in turn will increase the cost of delivery, block innovation, and divert resources that should be spent on producing better services for clinicians, commissioners, and patients.

There is currently substantial appetite for better use of data and software in the NHS [49]. This will only happen if the system engages constructively and technically with the individuals and teams who actually use that NHS data on a daily basis. We hope this paper will stimulate further dialogue between data providers and end users; we welcome feedback and further examples of both good and poor practice, and we are keen to engage, on both the details and broader strategic issues, with all members of the NHS and wider community.

Acknowledgments

The barriers described in this paper, and the work to overcome them, represent the work of the authors' whole technical and research team in the DataLab over the preceding 4 years. The authors are grateful for the hard work put in by all including the following: Richard Croker, Anna Powell-Smith, Dave Evans, Lisa French, Peter Inglesby, Alex Walker, Helen Curtis, and Brian MacKenna. They are also grateful for comments on an earlier draft from Jeni Tennison of the Open Data Institute, London, United Kingdom. Finally, the authors are grateful to those individuals from various NHS arm's length bodies who have helped their team understand and overcome some of the barriers documented. There was no specific funding for this paper. The authors' work on health care analytics is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre, Oxford; a Health Foundation grant (award reference number 7599); an NIHR School of Primary Care Research grant (award reference number 327). Funders had no role in the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Authors' Contributions

BG and SB conceived the paper. SB wrote the first draft. Both authors revised and approved the final manuscript. BG supervised the project and is guarantor.

Conflicts of Interest

All authors have completed the standard ICMJE uniform disclosure form and declare the following: BG has received research funding from the Laura and John Arnold Foundation, the Wellcome Trust, the Oxford Biomedical Research Centre, the NHS NIHR School of Primary Care Research, the Health Foundation, and the World Health Organization; he also receives personal income from speaking and writing for lay audiences on the misuse of science and is Chair of the Health Tech Advisory Board, reporting to the Secretary of State for Health and Social Care. SB is employed on BG's grants.

References

1. Open Knowledge Open Definition Group. Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. URL: http://opendefinition.org/od/2.1/en/ [accessed 2019-11-27]
2. Open Data White Paper | Unleashing the potential. UK: The Stationary Office Limited; Jun 2012.
3. NHS England: Annual report 2017/18. UK: APS Group; 2018.
4. NHS Digital. Transparency. URL: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/nhs-digital-s-annual-reports-and-accounts/nhs-digital-annual-report-and-accounts-2016-to-2017/department-of-health-hscic-framework-agreement/transparency [accessed 2019-11-27]
5. UK Statistics Authority. 2010 Oct. Statistics on Prescribing and Pharmaceutical Services in England: Assessment of Compliance With the Code of Practice for Official Statistics. URL: https://www.statisticsauthority.gov.uk/archive/assessment/assessment/assessment-reports/assessment-report-59---prescribing-and-pharmaceutical-services-in-england.pdf [accessed 2019-11-27]
6. Croker R, Walker AJ, Bacon S, Curtis HJ, French L, Goldacre B. New mechanism to identify cost savings in English NHS prescribing: minimising 'price per unit', a cross-sectional study. BMJ Open 2018 Feb 8;8(2):e019643. [doi: 10.1136/bmjopen-2017-019643] [Medline: 29439078]
7. Davidson P. Assets: how they work | GOV.UK Developer Documentation. 2010 Oct 5. Designing URI Sets for the UK Public Sector. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-URI-sets-uk-public-sector.pdf [accessed 2019-11-27]
8. Hyland B, Atemezing G, Villazón-Terrazas B. World Wide Web Consortium (W3C). 2014. Best Practices for Publishing Linked Data. URL: http://www.w3.org/TR/2014/NOTE-ld-bp-20140109/ [accessed 2019-11-27]

9. Welcome to GOV UK. 2012 Jun 28. Open Data: Unleashing the Potential. URL: https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential [accessed 2019-11-27]
10. NHS Digital Practice Level Prescribing Data. In: NHS Digital. [Internet]. Jun 2016 [cited Mar 2018]. URL: http://content.digital.nhs.uk/article/2021/Website-Search?q=presentation+level+data&go=Go&area=both
11. NHS Digital. Practice Level Prescribing Data. URL: https://digital.nhs.uk/data-and-information/publications/statistical/practice-level-prescribing-data [accessed 2019-11-27]
12. NHS England Data Catalogue. NHS England Data Catalogue: Prescribing. URL: https://data.england.nhs.uk/dataset?sort=score+desc%2C+metadata_modified+desc&q=&tags=prescibing&_tags_limit=0 [accessed 2019-11-27]
13. The NHS Business Services Authority (NHSBSA). Detailed Prescribing Information. URL: https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data/detailed-prescribing-information [accessed 2019-11-27]
14. The NHS Business Services Authority (NHSBSA). ISP Report Information. URL: https://www.nhsbsa.nhs.uk/information-services-portal-isp/isp-report-information [accessed 2019-11-27]
15. The NHS Business Services Authority (NHSBSA). Hints & Tips. URL: https://www.nhsbsa.nhs.uk/sites/default/files/2018-10/Information%20Services%20-%20Issue%2032%20%28V1%29%2010.2018.pdf [accessed 2019-11-27]
16. BNF Code Information. In: NHS BSA Information Services Portal. URL: https://apps.nhsbsa.nhs.uk/infosystems/data/showDataSelector.do?reportId=126
17. NHS Digital. 2018. Practice Level Prescribing - Glossary of Terms. URL: https://digital.nhs.uk/data-and-information/areas-of-interest/prescribing/practice-level-prescribing-in-england-a-summary/practice-level-prescribing-glossary-of-terms [accessed 2019-11-27]
18. Bacon S. Github. 2017. Understand BNF Code Publication Scheme. URL: https://github.com/ebmdatalab/openprescribing/issues/456 [accessed 2019-11-27]
19. Curtis H, French L, Bacon S. GitHub. 2017 Mar. Practices Where Setting = 4 Appears to Be Incorrect. URL: https://github.com/ebmdatalab/openprescribing/issues/323 [accessed 2019-11-27]
20. Inglesby P, Croker R, Bacon S. Github. 2018 Jul. Understand How Y03584's CCG Was Set. URL: https://github.com/ebmdatalab/openprescribing/issues/928 [accessed 2019-11-27]
21. Bacon S, French L, Croker R, Curtis H. GitHub. 2017 Feb. Deal With Small List Sizes. URL: https://github.com/ebmdatalab/openprescribing/issues/321 [accessed 2019-11-27]
22. NHS Business Services Authority. What is the Procedure for Setting Up a New Practice? URL: https://contactcentreservices.nhsbsa.nhs.uk/selfnhsukokb/AskUs_PS/en-gb/7601/gp-practices/43892/what-is-the-procedure-for-setting-up-a-new-practice [accessed 2019-11-27]
23. NHS England. CCG Map. URL: https://data.england.nhs.uk/dataset/ccg-map [accessed 2019-11-27]
24. Open Geography portal. Clinical Commissioning Groups (April 2017) Super Generalised Clipped Boundaries in England V4. URL: http://geoportal.statistics.gov.uk/datasets/clinical-commissioning-groups-april-2017-super-generalised-clipped-boundaries-in-england-v4 [accessed 2019-11-27]
25. Bacon S. GitHub. 2018 Jun. Handle Dormant and Closed Practices. URL: https://github.com/ebmdatalab/openprescribing/issues/891 [accessed 2019-11-27]
26. The NHS Business Services Authority (NHSBSA). Organisation and Prescriber Changes. URL: https://www.nhsbsa.nhs.uk/ccgs-area-teams-and-other-providers/organisation-and-prescriber-changes [accessed 2019-11-27]
27. Primary Care Support - NHS England. Practice Mergers and Closures Notifications. URL: https://pcse.england.nhs.uk/services/practice-mergers-and-closures-notifications/ [accessed 2019-11-27]
28. Bacon S. WhatDoTheyKnow. 2017. GP Practice Codes: Closures and Mergers Patient List Data. URL: https://www.whatdotheyknow.com/request/gp_practice_codes_closures_and_m [accessed 2019-11-27]
29. Bacon S. WhatDoTheyKnow. 2017. Practice Mergers and Closures Notifications Data. URL: https://www.whatdotheyknow.com/request/practice_mergers_and_closures_no [accessed 2019-11-27]
30. Bacon S. GitHub. 2018. Practice Merger Data Makes No Sense. URL: https://github.com/ebmdatalab/openprescribing/issues/932 [accessed 2019-11-27]
31. Bacon S. GitHub. 2018. What is the Unit of Measurement for 'quantity'? URL: https://github.com/ebmdatalab/openprescribing/issues/937 [accessed 2019-11-27]
32. The NHS Business Services Authority (NHSBSA). Prescription Cost Analysis (PCA) data. URL: https://www.nhsbsa.nhs.uk/prescription-data/dispensing-data/prescription-cost-analysis-pca-data [accessed 2019-11-27]
33. Bacon S. GitHub. 2017. When Does 'quantity' Represent Packs? URL: https://github.com/ebmdatalab/price-per-dose/issues/12 [accessed 2019-11-27]
34. WhatDoTheyKnow. Practice Manager's Email Addresses Within NHS England. URL: https://www.whatdotheyknow.com/request/practice_managers_email_addresse_193 [accessed 2019-11-27]
35. NHS England. How to make a Freedom of Information request to NHS England. URL: https://www.england.nhs.uk/contact-us/foi/ [accessed 2019-11-27]
36. French L. WhatDoTheyKnow. GP Practice Contact Information (FOI request to DHSC). URL: https://www.whatdotheyknow.com/request/gp_practice_contact_information [accessed 2019-11-27]

37. French L. WhatDoTheyKnow. GP Practice Contact Information (FOI request to MHRA). URL: https://www.whatdotheyknow.com/request/gp_practice_contact_information_5 [accessed 2019-11-27]
38. French L. WhatDoTheyKnow. GP Practice Contact Information (FOI request to NHS BSA). URL: https://www.whatdotheyknow.com/request/gp_practice_contact_information_4 [accessed 2019-11-27]
39. French L. WhatDoTheyKnow. GP Practice Contact Information (FOI request to NHS Digital). URL: https://www.whatdotheyknow.com/request/gp_practice_contact_information_3 [accessed 2019-11-27]
40. WhatDoTheyKnow. FOI requests by EBM DataLab to CCGs for practice email addresses. URL: https://www.whatdotheyknow.com/list/all?utf8=%E2%9C%93&query=%22Medicine+Optimisation+contact+information+for+GP+practices+in+your+CCG%22&request_date_after=&request_date_before=2017%2F12%2F01&variety=sent [accessed 2019-11-27]
41. The NHS Business Services Authority (NHSBSA). Drug Tariff Part VIII. URL: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff/drug-tariff-part-viii [accessed 2019-11-27]
42. Pharmaceutical Services Negotiating Committee. Price Concession and NCSO Archive. URL: https://psnc.org.uk/dispensing-supply/supply-chain/generic-shortages/ncso-archive/ [accessed 2019-11-27]
43. Stokes A. The NHS Business Services Authority (NHSBSA). FOI Information Request List Page. URL: https://apps.nhsbsa.nhs.uk/FOI/foiRequestList.do [accessed 2019-11-27]
44. Response to FOI request 6046 (NHSBSA). URL: https://apps.nhsbsa.nhs.uk/FOI/foiRequestDetail.do?bo_id=6046
45. The NHS Business Services Authority (NHSBSA). Response to FOI request 6808 (NHSBSA). URL: https://apps.nhsbsa.nhs.uk/FOI/foiRequestDetail.do?bo_id=6808 [accessed 2019-11-27]
46. Croker R. EBM DataLab. 2017 Dec 11. Getting the Data for the Drug Tariff Tool (*sigh*). URL: https://ebmdatalab.net/getting-the-data-for-the-drug-tariff-tool-sigh/ [accessed 2019-11-27]
47. NHS England Data Catalogue. URL: https://data.england.nhs.uk/ [accessed 2019-11-27]
48. GitHub. EBM Data Lab. URL: https://github.com/ebmdatalab [accessed 2019-11-27]
49. Department of Health and Social Care. Welcome to Gov UK. 2018 Oct 17. The Future of Healthcare: Our Vision for Digital, Data and Technology in Health and Care. URL: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care [accessed 2019-07-22]

Abbreviations

BNF: British National Formulary
CCG: clinical commissioning group
DHSC: Department of Health and Social Care
FOI: Freedom of Information
GP: general practitioner
NIHR: National Institute for Health Research
NHS: National Health Service
NHSBSA: National Health Service Business Services Authority
ONS: Office for National Statistics
PCSE: Primary Care Support England
PDPI: Practice Detailed Prescribing Information
RCT: randomized controlled trial
SNOMED: Systematized Nomenclature of Medicine
SQL: Structured Query Language

Edited by G Eysenbach; submitted 24.07.19; peer-reviewed by M Hawker, H Leroux; comments to author 01.09.19; revised version received 03.09.19; accepted 03.09.19; published 13.01.20.

Please cite as:
Bacon S, Goldacre B
Barriers to Working With National Health Service England's Open Data
J Med Internet Res 2020;22(1):e15603
URL: https://www.jmir.org/2020/1/e15603
doi:10.2196/15603
PMID:31929101

Viewpoint

The True Colours Remote Symptom Monitoring System: A Decade of Evolution

Sarah M Goodday[1,2*], PhD; Lauren Atkinson[1,3], MSc; Guy Goodwin[1,4], FMedSci; Kate Saunders[1,4], DPhil, MRCPsych, FHEA; Matthew South[1], PhD; Clare Mackay[1,4], PhD; Mike Denis[1], MSc; Chris Hinds[1,5], PhD; Mary-Jane Attenburrow[1,4*], MBBS, MRCPsych; Jim Davies[5,6], PhD; James Welch[5,7], BA (Oxon); William Stevens[6], PhD; Karen Mansfield[1], PhD; Juulia Suvilehto[1], PhD; John Geddes[1,4], MD, MRCPsych

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[2]4YouandMe, Seattle, WA, United States
[3]Oxford Center for Human Brain Activity, University of Oxford, Oxford, United Kingdom
[4]Oxford Health NHS Foundation Trust, Oxford, United Kingdom
[5]Big Data Institute, University of Oxford, Oxford, United Kingdom
[6]Department of Computer Science, University of Oxford, Oxford, United Kingdom
[7]NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Sarah M Goodday, PhD
Department of Psychiatry
University of Oxford
Warneford Lane
Oxford, OX3 7JX
United Kingdom
Phone: 44 (0)1865 618200
Email: sarah.goodday@psych.ox.ac.uk

Abstract

The True Colours remote mood monitoring system was developed over a decade ago by researchers, psychiatrists, and software engineers at the University of Oxford to allow patients to report on a range of symptoms via text messages, Web interfaces, or mobile phone apps. The system has evolved to encompass a wide range of measures, including psychiatric symptoms, quality of life, and medication. Patients are prompted to provide data according to an agreed personal schedule: weekly, daily, or at specific times during the day. The system has been applied across a number of different populations, for the reporting of mood, anxiety, substance use, eating and personality disorders, psychosis, self-harm, and inflammatory bowel disease, and it has shown good compliance. Over the past decade, there have been over 36,000 registered True Colours patients and participants in the United Kingdom, with more than 20 deployments of the system supporting clinical service and research delivery. The system has been adopted for routine clinical care in mental health services, supporting more than 3000 adult patients in secondary care, and 27,263 adolescent patients are currently registered within Oxfordshire and Buckinghamshire. The system has also proven to be an invaluable scientific resource as a platform for research into mood instability and as an electronic outcome measure in randomized controlled trials. This paper aimed to report on the existing applications of the system, setting out lessons learned, and to discuss the implications for tailored symptom monitoring, as well as the barriers to implementation at a larger scale.

(J Med Internet Res 2020;22(1):e15188) doi:10.2196/15188

KEYWORDS

symptom assessment; signs and symptoms; digital health; ecological momentary assessment; mood disorders

Introduction

The advancement of digital technology will gradually continue to shape how we measure, monitor, and manage health. A wide range of digital symptom monitoring tools exist, but there is a lack of evidence regarding their effectiveness in a health care context, particularly in the area of mental health. Such evidence will arise only from studies involving significant usage,

conducted in close partnership with clinicians, patients, and managers. For example, digital tools for patient-reported outcome measures (PROMs) are becoming standard practice in randomized controlled trials (RCTs) in many areas [1], and meta-analyses [2,3] have confirmed their equivalence with paper-based approaches.

True Colours is a digital tool, developed over a decade ago by psychiatrists, software engineers, and researchers at the University of Oxford, which has achieved significant usage. The initial version was used for remote monitoring of mood disorders, allowing patients and their clinicians to record and review symptom change. The recognized need to capture and monitor higher frequency phenotype information, particularly for conditions such as bipolar disorder (BD), is not new. Hard copy symptom monitoring diaries have been used for decades. However, these are limited by practicality issues.

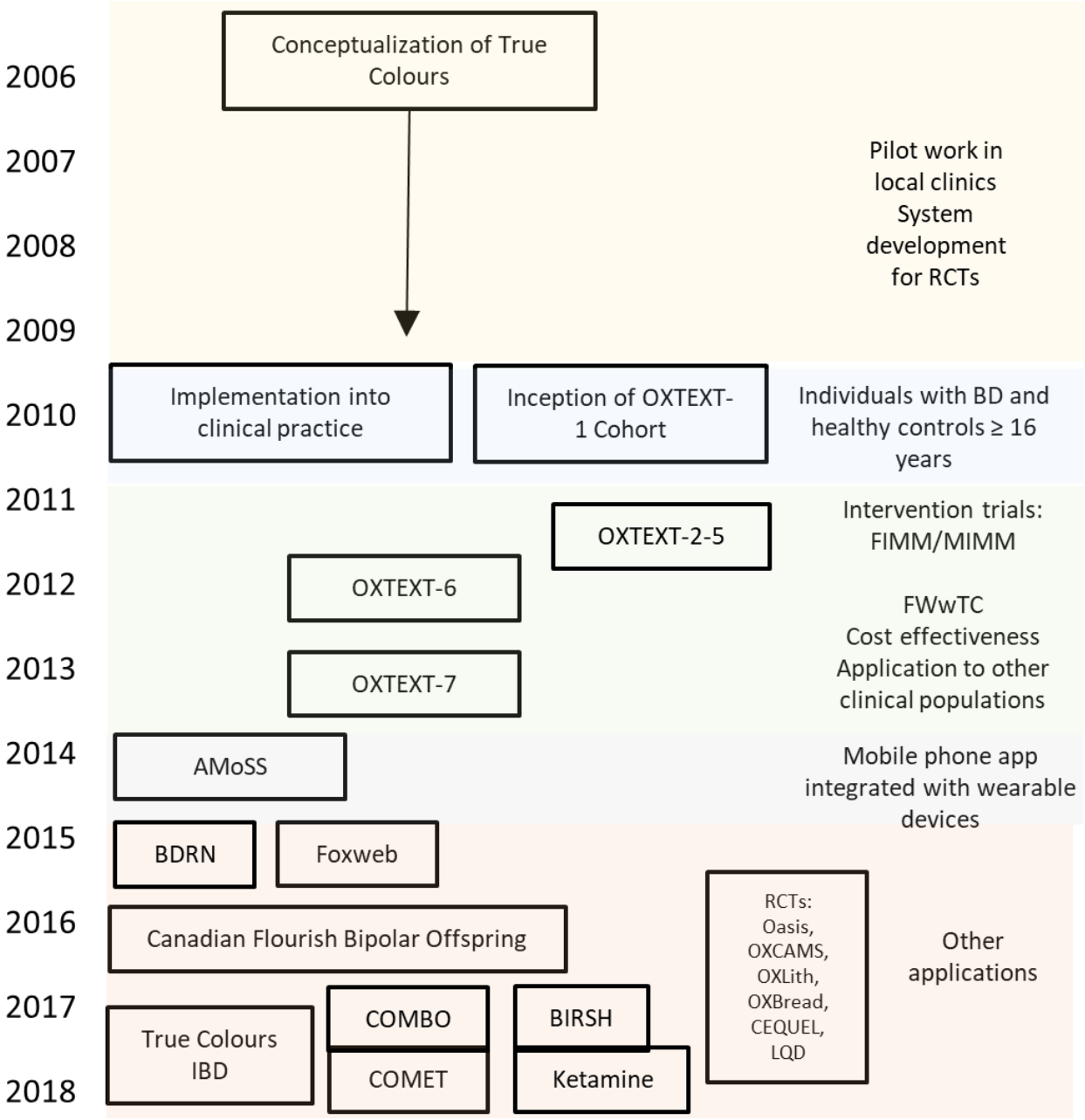
The True Colours system has many advantages over paper-based approaches toward the capturing of detailed, timed phenotype information, including the following: the ability to prompt for contemporaneous input, the automatic calculation of summary scores, the visualization of changes over time, and the provision of real time, as well as historical data to support clinical review, assessment, and early intervention. From a research perspective, the tool has additional advantages: eliminating errors in the transcription of information from paper forms, supporting a higher frequency of prompted, directed phenotyping, and reducing the recall bias associated with the recording of symptoms. Subsequent versions of the tool have added new functionality for data entry, patient or cohort management, and research delivery.

The system has been applied across several patient, participant, and high-risk populations, being used across 21 unique research and clinical service settings in the Oxfordshire and Buckinghamshire regions in the United Kingdom. Over the past decade, there have been over 36,000 registered True Colours participants from whom over 1.4 million questionnaire responses have been collected. Several feasibility studies and clinical service applications support the potential of True Colours as a larger scale symptom monitoring system, an electronic PROM, and a tool for digital phenotyping. This paper aimed to describe the evolution of the tool, its applications, and achievements and to discuss the potential for future wider application and integration.

Research Applications

The True Colours system was originally designed to monitor mood symptoms in adult patients with BD, attending the BD Research Clinic at the Department of Psychiatry at the University of Oxford, and it was designed for use in clinical trials, evolving from the Oxford University Symptom Monitoring System [4,5]. The original version of the system involved automated weekly prompts, delivered by text message or email (chosen by preference), for patients to complete self-reported measures of symptoms, including depression (16-item Quick Inventory of depressive symptoms) [6] and mania (5-item Altman Self Rating Mania Scale) [7], and other measures, such as anxiety (Generalized Anxiety Disorder Scale–7) [8], quality of life (EQ-5D) [9], and lifestyle behaviors. The system has expanded to include a wide range of symptoms from validated scales and bespoke measures tailored to specific research projects. As part of the True Colours platform, total symptom scores were presented graphically via a secure website and made available to patients, participants, and clinicians upon request. Over the past decade, the use of True Colours has expanded to several different research cohorts and patient populations (Figure 1).

Figure 1. Evolution and applications of True Colours. AMoSS: Automated Monitoring of Symptom Severity Study; BD: bipolar disorder; BDRN: Bipolar Disorder Research Network; CEQUEL: Comparative evaluation of quetiapine plus lamotrigine; COMBO: Collaborative Care Model for Bipolar Disorder; FIMM: Facilitated Integrated Mood Management; FWwTC: Feeling Well with True Colours; IBD: inflammatory bowel disease; LQD: lithium versus quetiapine augmentation for treatment resistant depression; MIMM: Manualized Integrated Mood Management; OxBREaD: Oxford Brain Body Research into Eating Disorders; OxCAMS: Oxford Study of Calcium Channel Antagonism, Cognition, Mood instability and Sleep; OxLith: Oxford Lithium Trial; RCT: randomized controlled trial.



The OXTEXT Program

Earlier work involving the University of Oxford Symptom Monitoring System established that technology-assisted symptom monitoring was acceptable to patients over a period of 36 weeks with 75% compliance [4], meaning that, on average, patients reported symptoms in response to prompts 75% of the time over follow-up. The OXTEXT program was dedicated to developing and validating the True Colours remote symptom monitoring system for patients, at a larger scale. Across several projects, the program revised the software after in-depth patient consultation, established a cohort of well-characterized patients with BD by using the improved system, determined the potential cost-effectiveness of this remote capture tool in clinical service, and tested remote mood monitoring as a potential intervention via RCTs (OXTEXT research studies 1-6). Several publications resulted from these studies, largely from the OXTEXT-1 cohort comprising up to 367 patients (≥16 years of age) from Oxfordshire, with a Diagnostic and Statistical Manual of Mental Disorders-IV diagnosis of BD (BDI, BDII, or BD-not otherwise specified), with some patients completing up to 81 months of continuous weekly mood measures. Compliance and acceptability of the True Colours system in the OXTEXT-1 cohort were excellent, with low attrition (<2%) and a median of less than 8% of weeks of missing data that did not differ by key sociodemographic factors or by mood score [10]. This pilot

work has demonstrated support for the feasibility of True Colours as a remote mood monitoring system in patients with mood disorders [4,10], and it has lent important insights into detectable mood instability that differentiates clinical course [5] and other BD patient characteristics, including cognitive functioning [10,11].

The OXTEXT-2 study [12] assessed participant's compliance with monitoring, their mental health resource use (including hospitalizations and face-to-face and phone contact with mental health staff), and service and medication costs, before and during their first 12 months of engagement with True Colours. Compliance with monitoring was high, with a median response completion rate of 92% for both Web-based and SMS symptom reporting and all patients continuing to report during the duration of the study. The introduction of True Colours was thought likely to reduce service costs, but this was not supported in OXTEXT-2. In fact, when associated with enhanced specialist care, medication costs increased over the first year of monitoring. This illustrated that studies of any digital addition to care need to account for all possible confounders relating to mood monitoring and mental health service costs. OXTEXT-2 did not examine nonmental health service costs, and larger economic evaluations of the True Colours system are required and are being conducted.

The True Colours system was also utilized as part of a psycho-education intervention for 121 patients with BD in an RCT (OXTEXT-6) [13]. The Facilitated Integrated Mood Management (FIMM) [14] study condition involved True Colours mood monitoring, a psycho-education manual, and individual sessions with a facilitator. This was compared with Manualized Integrated Mood Management, which only involved the psycho-education manual. Patients in the FIMM arm showed better knowledge of BD, and greater BD knowledge was associated with a high number of months in remission over 1-year follow-up [13]. Of note, True Colours in isolation is not intended as an intervention, but it may improve symptoms via insight into patients about their symptoms and closer, more accurate monitoring by clinicians, which will require further study.

The OXTEXT-7 study commenced in 2013 involving a trial rolled out to all 11 community mental health treatment services across Oxfordshire and Buckinghamshire titled as Feeling Well with True Colours (FWwTC). The goal of FWwTC was to offer patients a self-monitoring system that could allow care interventions to be tailored to the individual. Patients and clinicians create tailored symptom monitoring schedules on the basis of the type of symptom measure, frequency of prompts (weekly, daily, and several times a day), and reminder frequency. This study was a stepped-wedge, cluster randomized design. In this design, all services eventually implemented FWwTC, but the time at which they were trained to implement FWwTC was randomized to compare outcomes in treatment services before and after the introduction of FWwTC. The aim of this phase of OXTEXT was to apply True Colours to other patient populations (including those experiencing depression, anxiety, psychosis, alcohol and drug use, and BD) and test the feasibility and cost-effectiveness of such a tool in a larger scale secondary care setting. Experience from this trial is currently being synthesized, and it has proved heuristically useful [15], although uptake across clinical services was a challenge, illustrating the considerable barriers to innovation that persist in the National Health Service and other medical services.

Digital Phenotyping Studies

Digital phenotyping is the individual-level high-resolution data capture enabled by digital devices. The promise in this data capture is its ability to collect passive or active information in a real-world setting unbound to clinical visits. This affords the opportunity to discover new trajectories of signs and symptoms of disease, resulting in refined phenotypes and better detection and management of illness. The Collaborative Network for Bipolar Research to Improve Outcomes (ConBrio) [16] was a translational research program aimed at bringing together basic and clinician scientists in mathematics, computational biology, cognitive neuroscience, and neuroimaging. Central to the ConBrio program is the use of True Colours complemented by other methods for deep and frequent mood phenotyping to accelerate understanding and treatment of BD. This program has supported several projects, such as the Automated Monitoring of Symptom Severity Study (AMoSS), the use of True Colours in several RCTs, for example, Oxford Study of Calcium Channel Antagonism, Cognition, Mood instability and Sleep (OxCaMS) and Oxford Lithium Trial (OxLith), and other large phenotyping studies from the BD Research Network (BDRN) [17].

Automated Monitoring of Symptom Severity Study

Taking advantage of the developments in digital technology and ubiquity of mobile phones, the AMoSS study introduced a mobile phone app, Mood Zoom, to facilitate a higher frequency of symptom monitoring and included wearable devices as measures of objective symptoms. The Mood Zoom questionnaire comprises mood state descriptor items that are rated on a scale from 1 to 7 [18], which could be completed several times a day. Mood Zoom was used alongside weekly True Colours mood monitoring to help understand, in greater detail, mood episodes and mood instability in patients with BD and borderline personality disorder, as well as healthy volunteers in a sample of 139 patients with 3 months of continuous data (as per protocol) but with over 12 months of continuous data (for those willing to continue). The introduction of a mobile phone app also enabled the collection of passive background data, such as number of texts or calls and geolocation [19], which could reflect proxies of behavior associated with BD and how they are associated with mood, an emerging area with promise for the identification of behavioral markers of impending BD-related episodes [20]. Quantitative [18,21-23] studies have supported the feasibility and acceptability of the use of the Mood Zoom app and True Colours for daily and weekly symptom monitoring in patients with BD, borderline personality disorder, and controls. Specifically, attrition was low in the AMoSS cohort, with only 1 subject withdrawing and 8 subjects being excluded because of providing data for less than 2 months. Median adherence for the Mood Zoom and weekly measures was greater than 80% and 85%, respectively, and it remained stable over the study follow-up [18]. A qualitative study of 20 subjects from the AMoSS cohort provided support for the fact that reporting

on symptoms once daily was of no inconvenience, and it was felt that the system contributed to insights into personal symptoms and patterns [24]. Additional themes from this study highlight the importance of tailoring patient preferences into symptom reporting tools.

In recent studies, additional objective physiological measures, derived from Fitbit and wrist-worn accelerometers, were included, along with daily and weekly mood monitoring [22] as well as the proteus patch [21,25] that provides an estimate of heart rate. These studies have contributed insights into detectable variability of sleep patterns in patients with BD and borderline personality disorder, which map onto observable symptoms of low and irritable mood [21] and variability in mood [25]. The additional add-on of wearables offers an exciting line of inquiry into objective symptoms of illness-alleviating biases relating to subjective reporting of symptoms. This potentially supports downstream applications of True Colours, with the inclusion of additional devices for the measurement of objective symptoms, which will be important for deeper insights into early signs of disease.

Other Mood-Related Research Applications

The BDRN [17] adopted the True Colours system, engaging 815 research participants (815/4080, 19.97% of invited existing BDRN participants) with mood disorders [26]. BDRN participants with a diagnosis of BDII were more likely to register with True Colours. Approximately 78.2% (637/815) of registered participants completed 3 months of symptom reporting, approximately 51.1% (413/808) of the participants completed more than 1 year, and some participants continued mood monitoring for up to 3 years, demonstrating the feasibility of such a remote mood monitoring system at a larger scale.

An international application of True Colours is from the Canadian Flourish High-risk Offspring Study [27], recruiting young offspring of a parent with BD. The Flourish group has piloted the Web-based True Colours monitoring system in 50 high-risk offspring of a bipolar parent and 108 control offspring of psychiatrically well parents. Compliance was good over 30 days, with approximately 80% and greater than 90% of high-risk and control offspring completing daily ratings, respectively, and no difference in compliance between study groups. Daily mood scores significantly differentiated the high-risk from control offspring, and irregularity in weekly mood and anxiety scores was higher in high-risk offspring with remitted major mood disorders compared with those with no lifetime history of major mood disorders [27].

Additional studies from the University of Oxford have made use of the True Colours system to elucidate mood variability in BD, involving determining the different nonlinear time series processes of mood instability and analytic techniques for appropriately detecting it from high-frequency time series data [5,28], as well as its associations with mental imagery [29].

Application to Randomized Controlled Trials

RCTs of treatment efficacy in psychiatric disorders are expensive and lengthy, given the needed follow-up time for full Diagnostic and Statistical Manual of Mental Disorders threshold mood episodes to develop. Traditional endpoint assessments using paper and pencil questionnaires or clinician-rated diagnostic episodes also ignore clinically significant symptoms not meeting full diagnostic threshold between episodes [30] and cognitive dysfunction [31], which could be used to determine earlier and more proximal treatment effects. Several RCTs have used True Colours as both primary electronic outcome assessments and secondary higher frequency outcome measurements. For example, a 12-week double blind RCT (CEQUEL) [32] assessed combination therapy with quetiapine plus lamotrigine versus quetiapine monotherapy plus lamotrigine placebo on depressive symptoms in 266 patients (≥16 years) with BD, recruited across 27 different United Kingdom clinics. Another completed single blind RCT (OASIS) [33] of 3755 university students across the United Kingdom used True Colours to measure outcomes to determine the effectiveness of a Web-based cognitive behavioral therapy for insomnia and other psychiatric symptoms, including psychosis, mood, and anxiety.

Other mood-related applications of True Colours for outcome assessment in ongoing RCTs include the OxLith [34], aimed to compare lithium with placebo on mood instability in adult patients with BD; a trial assessing the clinical effectiveness and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression [35]; and OxCaMS [36], which aims to assess the impact of a calcium channel blocker on cognition and brain activity in adults with mood instability. Finally, the Oxford Brain Body Research into Eating Disorders study [37] involves a pilot trial to assess the safety, acceptability, and feasibility of deep brain stimulation in patients diagnosed with severe eating disorders. Other funded large trials involving the True Colours system under development include the Pramipexole Therapy in Treatment Resistant Depression and Bipolar Depression (PAX-D and PAX-BD) [38,39].

Expansion to Other Populations

As the research and clinical utility of True Colours became evident, it naturally branched out to other populations and research contexts. The Cognition and Mood Evolution across Time study is aimed at measuring cognition and brain activity in healthy participants with various levels of mood instability—a useful application of True Colours, with the inclusion of daily mood monitoring and cognitive tasks [40].

The True Colours system has also been modified for community outpatients, with a diagnosis of psychosis using forensic psychiatric services (FOXWEB risk violence tool). This research application involved the development of a Web-based violence risk monitoring tool for psychosis, which provides visual feedback of patient scores to clinicians to guide risk assessment [41], and this is being further piloted in inpatients.

The Brief Interventions for Self-Harm (BIRSH) clinic [42] has piloted True Colours for self-harm prevention in patients (13-65 years) presenting to accident and emergency departments. The aim of this ongoing research and service evaluation application was to determine the effectiveness of a new clinical service incorporating remote symptom monitoring to reduce self-harm repetition and health service costs.

The True Colours inflammatory bowel disease (IBD) group has expanded the True Colours schedule to include daily measures of ulcerative colitis and Crohn's disease symptoms, as well as fortnightly quality-of-life and other validated measures of disease activity. The initial aim of the True Colours IBD project was to develop and test the feasibility of a predictive index of IBD. A 6-month pilot in 66 patients supported the initial feasibility of this system, with 76% adherence rate for daily measures and 86% patient retention [43]. Further work has supported associations between daily IBD symptom measures and biological measures of disease activity [44], and this has facilitated the prediction of whether escalation of therapy or clinical investigation would be needed [45]. Qualitative findings from this work suggest that patients felt more in control and empowered by the True Colours IBD system [43].

Clinical Service Applications

Several of the noted research applications have evolved into the use of True Colours for a purely patient monitoring and/or clinician monitoring tool, despite little infrastructure and resources to do so. As of January 10, 2019, almost 3000 patients with any psychiatric condition and more than 700 clinicians have registered with True Colours in adult community mental health treatment service clinics across Oxfordshire and Buckinghamshire. The uniqueness of this application of remote monitoring of symptoms is in the individualized approach. This enables patients to choose, in consultation with their health care professional, how they would like to self-monitor, directly aligning from qualitative work suggesting the preference of flexibility and personalization in a symptom monitoring tool [24]. The system has also been taken up by child and adolescent mental health services across the Oxfordshire region, with 27,263 registered users.

True Colours IBD is a prime showcase of what True Colours could evolve into—an integrated platform for individualized patient and clinician monitoring of symptoms and quality-of-life outcomes, with the potential to predict when more symptoms are expected and prevent unnecessary clinic visits. With further validation, the implications this model could have for reducing health care costs and burden on individuals are extensive. Since September 2019, there are currently more than 750 registered IBD patients, within the John Radcliffe Hospital in Oxfordshire, using True Colours as a monitoring tool. True Colours has also been applied as a patient-reported outcome monitoring tool in clinical service clinics, testing the effectiveness of Ketamine as a therapy for treatment-resistant depression [46] and for self-harm risk assessment as an extension to the ongoing pilot work conducted by BIRSH [42]. Finally, the Collaborative Care Model for BD is an ongoing project aimed at testing the feasibility of True Colours in a primary care setting to understand perspectives of the True Colours system from both patients and clinicians. This project also aims to engage different services (primary and secondary care clinicians) in the collaborative treatment of patients through the sharing of True Colours symptom ratings.

Discussion

Over the past decade, True Colours has transformed from a simple text message prompt and reply system to a personalized Web-based symptom monitoring tool. This tool is now applied across a number of clinical populations and is integrated into several clinics as part of routine clinical care across the Oxfordshire and Buckinghamshire regions. A small team at the University of Oxford and the Big Data Institute has been supporting the continued use of True Colours and its application across a wide range of settings. Despite the relatively little resource that has been put into sustaining this system, its progress and scale, to date, are quite impressive, largely driven by small independent research grants.

The utility of True Colours as a research tool is unequivocal. The existing research involving this tool has contributed to considerable advancements in knowledge of mood instability and its correlates in mood and personality disorders, which would not have been possible with traditional aperiodic research or clinic assessments. The potential linkage of True Colours' patient-reported data to electronic medical records data currently available within United Kingdom–Clinical Record Interactive Search—a national research platform comprising deidentified electronic patient medical records—could yield a rich source of high-frequency phenotyping information for future research. This data linkage could provide continuous measures of patient-reported symptoms occurring in real time, which could be mapped onto hospital visits and acute episodes of illness. This could afford the opportunity to fill in the gaps between clinic visits and determine early subsyndromal phases of illness that could reflect targets for prevention of episode recurrence or worsening of symptoms—a substantial scientific and clinical resource.

In 2017, there were 325,000 mobile health apps available internationally, including lifestyle interventions, symptoms trackers, and personal coaches [47]. A vast majority of these tools are not evidence based, and their ability to accurately measure symptoms or feasibly engage patients is largely unknown [48,49]. Only about 25% of digital health app users continue using the app after 10 uses [50], indicating challenges with low retention. Furthermore, with the rapid turnaround of digital health apps, it is difficult to rigorously test their effectiveness or implement into practice before they become obsolete [51]. Other symptom monitoring platforms include the Chrono-record [52], a computer-based symptom monitoring system, and the MONitoring treatment and pRediCtion BD episode system [53], an Android-based mobile phone objective and subjective symptom monitoring system designed for patients with BD. Patientslikeme [54] is a digital health platform in the United Kingdom, which involves a Web-based system that enables patients to track symptoms and view other members' health information. The *Patientslikeme* platform currently has 600,000 registered users, and it is meant to produce data for research purposes and provide empowerment and community to patients to track their own symptoms. These tools are useful in unique ways, but these are yet to have any integration with clinical service. In addition, they are targeted toward specific

conditions or the broad reporting of symptoms, some untethered to validated measures.

In an era where the digital health market is becoming increasingly saturated, careful integration of these tools within the health care system is crucial [55]. There is a need to develop digital remote monitoring tools that are evidence based [56], with infrastructure to support secure and sensitive personal information and enable the growth of the tool in tandem with rapidly developing digital technologies. Obvious barriers to this potential integration surround buy-in from health care providers, the potential to create inefficiencies, and data security concerns. This underscores the needed infrastructure for such a remote monitoring tool in clinical practice, with education for clinicians on its purpose and use, an electronic system with ease of access, and the flexibility and support to tailor the service to different patient populations and clinical care contexts. Uptake within clinical service will be a challenge and will require support from several participating parties.

What is unique about True Colours is the pilot work behind the tool's feasibility across different patient populations, and its use alongside clinical judgement. Its evolution has been guided by several feasibility studies, clinical and software development expertise, and, most importantly, participant, patient, and clinician feedback. The concept of True Colours as an integrated clinical care model offers benefits to patients through the returning of simple, visually effective symptom summaries, empowering individuals to play an active role in their health, which alone could have a therapeutic effect, as seen in other areas of medicine, such as oncology [57,58]. For clinical practice, this tool could enable clinicians to have access to continuous health information from their patients unbound to clinic visits, providing PROMs at higher frequencies and lending insight into dynamic fluctuations in symptoms that cannot be captured by traditional health measurement systems by self-report measures recalling symptoms over long periods of time. In turn, this could support real-time assessment and management of chronic conditions while freeing up time and resources for the National Health Service.

Acknowledgments

This work was funded by a Medical Research Council Mental Health Data Pathfinder Award. Additional funding for the True Colours system has come from the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Center. This funding body had no role in the design, writing, or interpretation of this paper. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, or Department of Health.

Conflicts of Interest

JRG reports grants from United Kingdom Medical Research Council, grants from Wellcome, grants from NIHR, outside the submitted work; JRG led the conception of True Colours, a digital phenotyping and outcome assessment tool, and JRG has overseen its implementation in routine clinical practice and research studies. He is also an NIHR Senior Investigator and Director of the NIHR Oxford Health Biomedical Research Centre. GG is an NIHR Emeritus Senior Investigator, holding shares in P1Vital and P1Vital products, and has served as consultant, advisor, or CME speaker in the last 3 years for Allergan, Angelini, Compass pathways, MSD, Janssen, Lundbeck (/Otsuka or /Takeda), Medscape, Minerva, P1Vital, Pfizer, Sage, Servier, Shire, and Sun Pharma. All other authors report no conflict of interest related to this paper.

References

1. Byrom B, Gwaltney C, Slagle A, Gnanasakthy A, Muehlhausen W. Measurement Equivalence of Patient-Reported Outcome Measures Migrated to Electronic Formats: A Review of Evidence and Recommendations for Clinical Trials and Bring Your Own Device. Ther Innov Regul Sci 2019 Jul;53(4):426-430. [doi: 10.1177/2168479018793369] [Medline: 30157687]
2. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 7;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
3. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
4. Bopp J, Miklowitz D, Goodwin G, Stevens W, Rendell J, Geddes J. The longitudinal course of bipolar disorder as revealed through weekly text messaging: a feasibility study. Bipolar Disord 2010 May;12(3):327-334 [FREE Full text] [doi: 10.1111/j.1399-5618.2010.00807.x] [Medline: 20565440]
5. Bonsall MB, Wallace-Hadrill SM, Geddes JR, Goodwin GM, Holmes EA. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. Proc Biol Sci 2012 Mar 7;279(1730):916-924 [FREE Full text] [doi: 10.1098/rspb.2011.1246] [Medline: 21849316]
6. Rush A, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 1;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]

7. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry 1997 Nov 15;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
8. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
9. 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]
10. McKnight RF, Bilderbeck AC, Miklowitz DJ, Hinds C, Goodwin GM, Geddes JR. Longitudinal mood monitoring in bipolar disorder: course of illness as revealed through a short messaging service. J Affect Disord 2017 Dec 1;223:139-145. [doi: 10.1016/j.jad.2017.07.029] [Medline: 28753472]
11. Bilderbeck AC, Reed ZE, McMahon HC, Atkinson LZ, Price J, Geddes JR, et al. Associations between mood instability and emotional processing in a large cohort of bipolar patients. Psychol Med 2016 Nov;46(15):3151-3160. [doi: 10.1017/S003329171600180X] [Medline: 27572660]
12. Simon J, Budge K, Price J, Goodwin G, Geddes J. Remote mood monitoring for adults with bipolar disorder: an explorative study of compliance and impact on mental health service use and costs. Eur Psychiatry 2017 Sep;45:14-19 [FREE Full text] [doi: 10.1016/j.eurpsy.2017.06.007] [Medline: 28728090]
13. Bilderbeck AC, Atkinson LZ, McMahon HC, Voysey M, Simon J, Price J, et al. Psychoeducation and online mood tracking for patients with bipolar disorder: a randomised controlled trial. J Affect Disord 2016 Nov 15;205:245-251. [doi: 10.1016/j.jad.2016.06.064] [Medline: 27454410]
14. Miklowitz DJ, Price J, Holmes E, Rendell J, Bell S, Budge K, et al. Facilitated Integrated Mood Management for adults with bipolar disorder. Bipolar Disord 2012 Mar;14(2):185-197 [FREE Full text] [doi: 10.1111/j.1399-5618.2012.00998.x] [Medline: 22420594]
15. Nickless A, Voysey M, Geddes J, Yu L, Fanshawe TR. Mixed effects approach to the analysis of the stepped wedge cluster randomised trial-Investigating the confounding effect of time through simulation. PLoS One 2018;13(12):e0208876 [FREE Full text] [doi: 10.1371/journal.pone.0208876] [Medline: 30543671]
16. Dimensions. CON BRIO: Collaborative Network for Bipolar Research to Improve Outcomes. URL: https://app.dimensions.ai/details/grant/grant.3639647 [accessed 2019-11-20]
17. Bipolar Disorder Research Network (BDRN). URL: http://www.bdrn.org [accessed 2019-11-20]
18. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Osipov M, Clifford G, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. J Affect Disord 2016 Nov 15;205:225-233 [FREE Full text] [doi: 10.1016/j.jad.2016.06.065] [Medline: 27449555]
19. 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 Aug;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
20. Goodday SM, Cipriani A. Challenges in identifying behavioural markers of bipolar disorder through objective smartphone data. Aust N Z J Psychiatry 2019 Feb;53(2):168-169. [doi: 10.1177/0004867418816813] [Medline: 30518223]
21. Carr O, Saunders KE, Bilderbeck AC, Tsanas A, Palmius N, Geddes JR, et al. Desynchronization of diurnal rhythms in bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Apr 12;8(1):79 [FREE Full text] [doi: 10.1038/s41398-018-0125-7] [Medline: 29643339]
22. Palmius NO, Bilderbeck A, Goodwin G, Saunders K, Tsanas A, Clifford G. A Multi-sensor Monitoring System for Objective Mental Health Management in Resource Constrained Environments. In: Proceedings of the Appropriate Healthcare Technologies for Low Resource Settings. 2014 Presented at: AHT'14; September 17-18, 2014; London, UK. [doi: 10.1049/cp.2014.0764]
23. Arribas IP, Goodwin GM, Geddes JR, Lyons T, Saunders KE. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Dec 13;8(1):274 [FREE Full text] [doi: 10.1038/s41398-018-0334-0] [Medline: 30546013]
24. Saunders K, Bilderbeck A, Panchal P, Atkinson L, Geddes J, Goodwin G. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
25. 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 Jan 26;8(1):1649 [FREE Full text] [doi: 10.1038/s41598-018-19888-9] [Medline: 29374207]
26. Gordon-Smith K, Saunders K, Geddes JR, Harrison PJ, Hinds C, Craddock N, et al. Large-scale roll out of electronic longitudinal mood-monitoring for research in affective disorders: report from the UK bipolar disorder research network. J Affect Disord 2019 Mar 1;246:789-793. [doi: 10.1016/j.jad.2018.12.099] [Medline: 30623825]
27. Duffy A, Keown-Stoneman CD, Goodday SM, Saunders K, Horrocks J, Grof P, et al. Daily and weekly mood ratings using a remote capture method in high-risk offspring of bipolar parents: compliance and symptom monitoring. Bipolar Disord 2019 Mar;21(2):159-167. [doi: 10.1111/bdi.12721] [Medline: 30422376]
28. Moore PJ, Little MA, McSharry PE, Goodwin GM, Geddes JR. Mood dynamics in bipolar disorder. Int J Bipolar Disord 2014 Dec;2(1):11 [FREE Full text] [doi: 10.1186/s40345-014-0011-z] [Medline: 26092397]

29. Holmes EA, Deeprose C, Fairburn CG, Wallace-Hadrill SM, Bonsall MB, Geddes JR, et al. Mood stability versus mood instability in bipolar disorder: a possible role for emotional mental imagery. Behav Res Ther 2011 Oct;49(10):707-713 [FREE Full text] [doi: 10.1016/j.brat.2011.06.008] [Medline: 21798515]
30. Harrison PJ, Geddes JR, Tunbridge EM. The emerging neurobiology of bipolar disorder. Trends Neurosci 2018 Jan;41(1):18-30 [FREE Full text] [doi: 10.1016/j.tins.2017.10.006] [Medline: 29169634]
31. Bourne C, Aydemir Ö, Balanzá-Martínez V, Bora E, Brissos S, Cavanagh JT, et al. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr Scand 2013 Sep;128(3):149-162. [doi: 10.1111/acps.12133] [Medline: 23617548]
32. Geddes JR, Gardiner A, Rendell J, Voysey M, Tunbridge E, Hinds C, CEQUEL Investigators and Collaborators. Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2 × 2 factorial randomised trial. Lancet Psychiatry 2016 Jan;3(1):31-39 [FREE Full text] [doi: 10.1016/S2215-0366(15)00450-2] [Medline: 26687300]
33. Freeman D, Sheaves B, Goodwin GM, Yu L, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
34. Saunders KE, Cipriani A, Rendell J, Attenburrow M, Nelissen N, Bilderbeck AC, et al. Oxford Lithium Trial (OxLith) of the early affective, cognitive, neural and biochemical effects of lithium carbonate in bipolar disorder: study protocol for a randomised controlled trial. Trials 2016 Mar 2;17(1):116 [FREE Full text] [doi: 10.1186/s13063-016-1230-7] [Medline: 26936776]
35. Marwood L, Taylor R, Goldsmith K, Romeo R, Holland R, Pickles A, et al. Study protocol for a randomised pragmatic trial comparing the clinical and cost effectiveness of lithium and quetiapine augmentation in treatment resistant depression (the LQD study). BMC Psychiatry 2017 Jun 26;17(1):231 [FREE Full text] [doi: 10.1186/s12888-017-1393-0] [Medline: 28651526]
36. Atkinson LZ, Colbourne L, Smith A, Harmer CH, Nobre AC, Rendell J, et al. The Oxford study of Calcium channel Antagonism, Cognition, Mood instability and Sleep (OxCaMS): study protocol for a randomised controlled, experimental medicine study. Trials 2019 Feb 12;20(1):120 [FREE Full text] [doi: 10.1186/s13063-019-3175-0] [Medline: 30755265]
37. Park RJ, Scaife JC, Aziz TZ. Study protocol: using Deep-Brain stimulation, multimodal neuroimaging and neuroethics to understand and treat severe enduring anorexia nervosa. Front Psychiatry 2018;9:24 [FREE Full text] [doi: 10.3389/fpsyt.2018.00024] [Medline: 29681866]
38. ISRCTN Registry. Is Pramipexole Effective as an Add-on Treatment for People With Treatment-resistant Depression? URL: http://www.isrctn.com/ISRCTN84666271 [accessed 2019-11-20]
39. Northern Centre for Mood Disorders. URL: https://mood-disorders.co.uk/ [accessed 2019-11-20]
40. Panchal P, Scholl J, Nelissen N, Saunders K, Darby D, Rushworth M, et al. Mood instability and reward processing: daily remote monitoring as a modern phenotyping tool for bipolar disorder. Eur Neuropsychopharmacol 2018;28(Suppl 1):S88-S89. [doi: 10.1016/j.euroneuro.2017.12.121]
41. Gulati G, Cornish R, Al-Taiar H, Miller C, Khosla V, Hinds C, et al. Web-based violence risk monitoring tool in psychoses: pilot study in community forensic patients. J Forensic Psychol Pract 2016 Jan 1;16(1):49-59 [FREE Full text] [doi: 10.1080/15228932.2016.1128301] [Medline: 26924945]
42. Brand F. FoNS Centre for Nursing Innovation. 2015. Developing, Implementing and Evaluating a Model for a Self Harm Clinic. URL: https://www.fons.org/Resources/Documents/Project%20Reports/BIRSH-REPORT-June-2015-.pdf [accessed 2019-11-20]
43. Walsh A. Semantic Scholar.: University of Oxford; 2017. TrueColours: Real Time Data Collection in Patients With Ulcerative Colitis. URL: https://pdfs.semanticscholar.org/32cc/acc043c3060f8c404efdaaaa718158d11f45.pdf [accessed 2019-11-20]
44. Probert F, Walsh A, Jagielowicz M, Yeo T, Claridge T, Simmons A, et al. Plasma nuclear magnetic resonance metabolomics discriminates between high and low endoscopic activity and predicts progression in a prospective cohort of patients with ulcerative colitis. J Crohns Colitis 2018 Nov 15;12(11):1326-1337 [FREE Full text] [doi: 10.1093/ecco-jcc/jjy101] [Medline: 30016408]
45. Walsh A, Matini L, Hinds C, Sexton V, Brain O, Keshav S, et al. Real-time data monitoring for ulcerative colitis: patient perception and qualitative analysis. Intest Res 2019 Jul;17(3):365-374 [FREE Full text] [doi: 10.5217/ir.2018.00173] [Medline: 31146510]
46. Oxford Health NHS Foundation Trust. Treatment. URL: https://www.oxfordhealth.nhs.uk/ketamine-service/treatment/ [accessed 2019-11-25]
47. Research 2 Guidance. 2017. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-11-20]
48. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]

49. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
50. Mclean V. PRWeb. 2011 Apr 25. Motivating Patients to Use Smartphone Health Apps. URL: http://www.prweb.com/releases/2011/04/prweb5268884.htm [accessed 2019-11-20]
51. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
52. Bauer M, Grof P, Gyulai L, Rasgon N, Glenn T, Whybrow PC. Using technology to improve longitudinal studies: self-reporting with ChronoRecord in bipolar disorder. Bipolar Disord 2004 Feb;6(1):67-74. [doi: 10.1046/j.1399-5618.2003.00085.x] [Medline: 14996143]
53. Faurholt-Jepsen M, Vinberg M, Christensen EM, Frost M, Bardram J, Kessing LV. Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder--the MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolAr disorder episodes): a randomised controlled single-blind trial. BMJ Open 2013;3(7):pii: e003353 [FREE Full text] [doi: 10.1136/bmjopen-2013-003353] [Medline: 23883891]
54. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2019-11-20]
55. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. Int J Bipolar Disord 2017 Dec;5(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0073-9] [Medline: 28155206]
56. Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 7;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
57. Nipp RD, El-Jawahri A, Ruddy M, Fuh C, Temel B, D'Arpino SM, et al. Pilot randomized trial of an electronic symptom monitoring intervention for hospitalized patients with cancer. Ann Oncol 2019 Feb 1;30(2):274-280. [doi: 10.1093/annonc/mdy488] [Medline: 30395144]
58. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]

Abbreviations

AMoSS: Automated Monitoring of Symptom Severity Study
BD: bipolar disorder
BDRN: BD Research Network
BIRSH: Brief Interventions for Self-Harm
ConBrio: Collaborative Network for Bipolar Research to Improve Outcomes
FIMM: Facilitated Integrated Mood Management
FWwTC: Feeling Well with True Colours
IBD: inflammatory bowel disease
NIHR: National Institute for Health Research
OxCaMS: Oxford Study of Calcium Channel Antagonism, Cognition, Mood instability and Sleep
OxLith: Oxford Lithium Trial
PROM: patient-reported outcome measure
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.07.19; peer-reviewed by A Tsolaki, D Gustafson; comments to author 14.09.19; revised version received 25.09.19; accepted 22.10.19; published 15.01.20.

Please cite as:
Goodday SM, Atkinson L, Goodwin G, Saunders K, South M, Mackay C, Denis M, Hinds C, Attenburrow MJ, Davies J, Welch J, Stevens W, Mansfield K, Suvilehto J, Geddes J
The True Colours Remote Symptom Monitoring System: A Decade of Evolution
J Med Internet Res 2020;22(1):e15188
URL: https://www.jmir.org/2020/1/e15188
doi:10.2196/15188
PMID:31939746

Viewpoint

A Stimulated Recall Method for the Improved Assessment of Quantity and Quality of Social Media Use

Nastasia Griffioen[1], MSc; Marieke M J W Van Rooij[1], PhD; Anna Lichtwarck-Aschoff[1], PhD; Isabela Granic[1], PhD

Developmental Psychopathology and Treatment, Radboud University Nijmegen, Nijmegen, Netherlands

Corresponding Author:
Nastasia Griffioen, MSc
Developmental Psychopathology and Treatment
Radboud University Nijmegen
Montessorilaan 3
Nijmegen
Netherlands
Phone: 31 024 3611614
Email: n.griffioen@bsi.ru.nl

Abstract

Background: Social media are as popular as ever, and concerns regarding the effects of social media use on adolescent well-being and mental health have sparked many scientific studies into use effects. Social media research is currently at an important crossroads: conflicting results on social media use's effects on well-being are abundant, and recent work in the field suggests that a new approach is required. The field is in need of an approach involving objective data regarding use where necessary and attention to different kinds of detail such as the why and how of social media use.

Objective: We present a novel paradigm implementing a principle from educational sciences called stimulated recall and demonstrate how it can be applied to social media use research. Our stimulated recall paradigm implements a number of elements that can fill the gaps currently present in social media and well-being research.

Methods: Objective data are collected regarding users' social media behaviors through video footage and in-phone data and used for a structured stimulated recall interview to facilitate detailed and context-sensitive processing of these objective data. In this interview, objective data are reviewed with the participant in an act of co-research, in which details such as the reasons for their use (eg, boredom) and processes surrounding their use (eg, with whom) are discussed and visualized in a stimulated recall chart.

Results: Our ongoing study (N=53) implementing this paradigm suggests this method is experienced as pleasant by participants in spite of its personal and intensive nature.

Conclusions: The stimulated recall paradigm offers interesting and necessary avenues for approaching social media use research from new angles, addressing aspects of use that have thus far remained underexposed. The answers to questions such as "Why do adolescents use social media?" "In what ways exactly do they use social media?" and "How does social media use make them feel in the moment?" are now within reach, an important step forward in the field of social media use and well-being research.

(J Med Internet Res 2020;22(1):e15529) doi:10.2196/15529

KEYWORDS

technology use; stimulated recall; social media; well-being; qualitative research; interview; digital technologies

Introduction

Digital technologies such as social media have seen an immense increase in adoption and popularity. Whereas in 2005 only 10% of the United States population reported using one or more social networking websites, in 2015 this percentage had skyrocketed to 65% for the entire population and to 90% for people aged 18 to 29 years [1]. Social media enable people to be more easily connected to others all around the globe, and their potential for expansion of social networks is likely what drives these platforms' popularity. Social media feed directly into the fundamental human need for social connection, which may be especially true for children and adolescents, who have grown up in a world in which digital technologies permeate almost every aspect of their daily lives (eg, in playing games, at school, in doing homework, chatting with friends, and even dating).

This synthesis of online and offline experiences in the lives of many children and adolescents has sparked a lot of debate among the general public as well as with researchers. Concerns about digital technologies center around screen time, since most technologies are accessed through screens that are carried around everywhere. Strong concerns have been raised regarding the effects of these screens, and social media in particular, on the well-being of youth [2,3]. As a result, many studies have been conducted to clarify (not without pressure from the lay public) what exactly the use of digital technologies is doing to youngsters' mental well-being and development [4-10]. The debate keeps raging on, and more and more studies are added to the already large body of work on the relationship between social media use and youth well-being. Yet there is strikingly little consensus on the matter, as illustrated by two recent literature reviews [11,12]: some studies indicate a negative relationship between social media use and well-being [6,13] and others a positive relationship [5,14,15].

This lack of unanimity in the field may have to do with important methodological limitations. First, social media use and well-being research has been largely characterized by a focus on quantity, operationalized by metrics like frequency and duration [5,10,15-19]. This is problematic because such metrics do not tell us anything about the types of activities, the contexts in which they take place, and how they are experienced by users. These types of context specifics, however, seem to be what differentiates negative and positive outcomes of social media use; for instance, whether social media are used actively or passively makes a difference to users' well-being [20]. Second, in most cases such metrics are being assessed using a method that is not particularly suited for these target variables—self-report [9,16,21-27]. Alarmingly enough, studies have shown that people are in fact notoriously bad at recalling details about their use of social media or other digital technologies [28-30]. If metrics such as duration and frequency of use are what we are interested in relative to well-being, it is vital that reliable, objective data on these behaviors are gathered rather than self-report data. Third, when self-report is used, it is generally in the context of observational studies, where no manipulation takes place [31-39], making it impossible to draw a causal inference. Additionally, when experimental designs are used, they mostly involve fabricated social media–like environments [40,41] rather than users' personal accounts, which offer much more salient and ecologically valid contexts for studies. Also, most experimental studies in the field arbitrarily choose one type of social media platform [7,20,42-46] at the exclusion of others, often meaning outdated apps are being studied, or only one app, when in fact youth use several simultaneously. Focusing on one platform also brings forth the danger of selection bias, since there may be differences (eg, age) between user bases of different platforms that can be relevant for a study and its outcomes.

Social media might be a context that requires a radically different approach, a new methodological lens—one that is objective and accurate, while considering the unique (ie, socially salient) digital context. Thus, to extend current research on social media use and address the pitfalls present (ie, use of retrospective self-report and a focus on quantity only), we suggest that a new approach should implement objective data where quantitative measures are concerned and include a context-sensitive aspect in which attention is paid to what users are doing exactly, who they interact with, and how these specific conditions and experiences make them feel. The functions (ie, why youth use social media) of and processes (ie, in what ways, with whom, and when youth use social media) surrounding social media use have simply not been addressed by the majority of studies in psychological science. Such research questions require an ecologically valid and detailed approach that allows for quantitative and qualitative data sources, and we suggest that stimulated recall holds promise in this area.

In his original version of the stimulated recall method [47], Bloom [48] played audio from lectures versus study discussions to his students and asked them to comment on their thoughts during these events in an attempt to investigate differences in learning processes between these two forms of teaching. According to Bloom, the primary aim of the method is "that the subject may be enabled to relive an original situation with vividness and accuracy if he is presented with a large number of the cues or stimuli which occurred during the original situation." As such, stimulated recall offers a way of investigating situations as they occur in the real world, without external influences or restraints. The method consists of two primary elements: one or multiple sources of objective information to aid the participant in recall and a qualitative, detailed interview of the participant's recall of the event of interest. This combination of quantitative and qualitative techniques seems to be exactly the sort of approach from which the field of social media use and well-being research could benefit. The collection of objective data helps address the current unreliability of measures while the in-depth investigation of users' activities, motives, and feelings helps to provide the detail and nuance that seems important. This new approach will ideally allow us to answer questions that are as of yet out of reach:

- Why do adolescents use social media in the first place?
- Which kinds of interactions do they experience on social media and with whom?
- What do adolescents expect from social media?
- How do these experiences make them feel?

Having discussed the origins and basics of stimulated recall, we will now present the methodology as it can be applied to social media use in young people, drawing examples from our own ongoing effort to implement this methodology in our study of social media use and well-being. This is an active study (started in April 2019) currently being conducted at the Radboud University Nijmegen, the Netherlands. Participants are students aged 18 to 25 years (N=53; 42 female) and are tested in the Bar Lab of the Behavioural Science Institute to ensure an informal atmosphere, predisposing participants to behave as they would in other public spaces rather than in a regular lab. The study was approved by the institution's Ethics Committee Social Sciences, approval number ECSW-2019-020.

Stimulated Recall for Social Media Research

Objective Data Sources

A key criterion for a successful implementation of stimulated recall is access to objective data or, more accurately, data that anchor the recall to directly observable behavior. This could take the form of audio data; videotaped recordings; screen captures of activity on a computer, game console, or phone; back-end data from games or apps that log activity; and so on. These are the data collected to scaffold the subsequent interview process and provide the necessary memory. Two important data sources for a social media research application of this paradigm, video footage and in-app information, will now be illustrated using elements of our ongoing study.

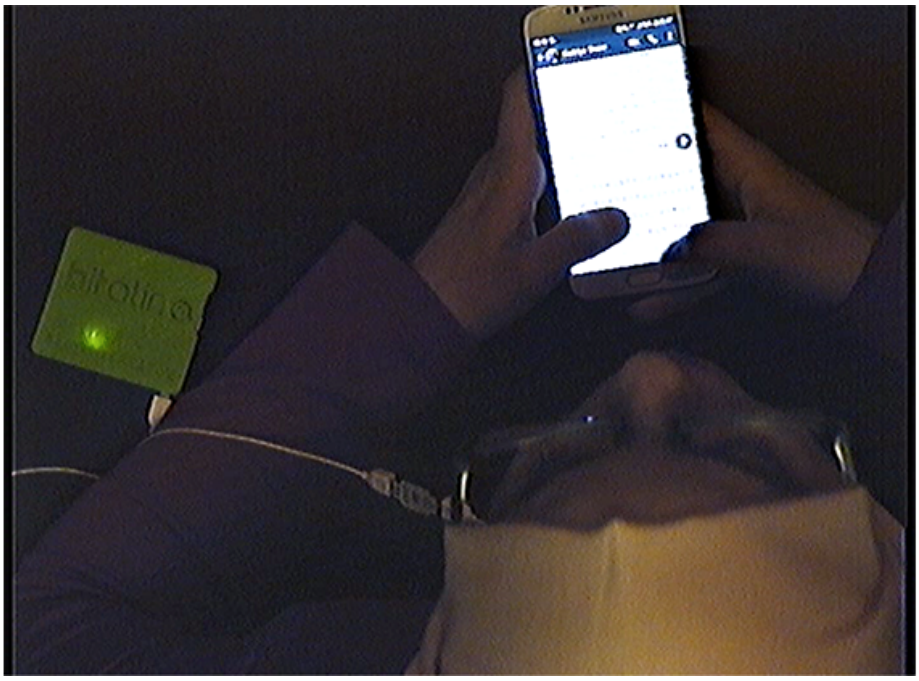
Video Recording

To enable a naturalistic capturing of student social media behaviors, participants in our study were asked to wait for 10 minutes after having completed a task. The details of the procedure prior to the waiting period will not be elaborated on here, but we would like to note that for half of the participants it included a stress manipulation in the form of the Leiden Public Speaking Task [49]. After having completed the first phase, participants were told that "in no more than 10 minutes" the researcher would return and the study would proceed as planned. During these 10 minutes, and unbeknownst to the participants, their activities were recorded using a video camera in the hopes of capturing naturalistic social media use.

Whether we would be able to capture smartphone behaviors of interest depended to a large extent on the camera setup. The best solution ultimately involved a camera installed right above the participant's seat, which has provided us with good and reliable footage (ie, the participant could change poses, but this would affect the quality of footage only minimally) of the participant's phone in all of the cases so far. Participants were always tested in the Bar Lab room seated at a table positioned directly under the camera. To ensure that participants would not get up and walk around the room (and thus leave the camera's field of view), they were asked to remain seated while the researcher was gone to ensure a steady signal from the physiological equipment (which, in reality, was robust to movement).

The dome-shaped camera was able to rotate on its axis, tilt, and zoom in and out, as well as adjust focus to points nearer or further away in space, allowing us to sharpen or blur the image as necessary. The camera was controlled by the researcher from a control room next to the study room where the participant was waiting. The focus point of the camera could be controlled in such a way that the participant's smartphone screen (if used by the participant) was visible but no text could be read from the screen to guarantee the privacy of the participant and any people whose information may have been featured on the screen. Similarly, images were always blurry, and, although shapes could be made out, any people featured on the participant's screen could not be identified. What these recordings did enable us to see, however, was which apps the participant was using and what the participant was doing in these apps (eg, just scrolling, typing text, liking a post). See Figure 1 for a screenshot of a pilot participant's recording. On this screenshot, for instance, we can see that the participant seems to be typing a message in WhatsApp, judging by the layout of the app visible on the screen. No preprocessing of the video footage is required before use in the interview, meaning that the interview can take place almost directly following the monitoring/waiting phase.

Figure 1. Screenshot from the video recording of one of our pilot participants.

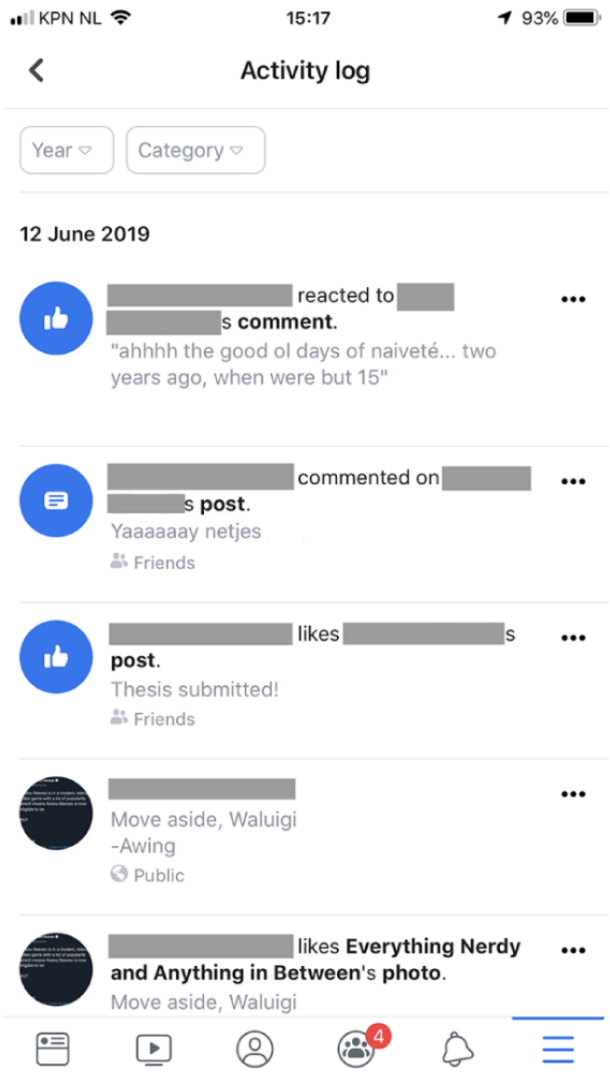


In-Phone Information

While the video recordings provide valuable information about participant activities, in a number of cases the recordings alone were not sufficient to capture our desired level of detail regarding participant phone use. For instance, very brief actions such as hitting a Like button could sometimes be harder to identify with certainty given the blurriness of the video image. In other cases, the layouts of apps were sometimes similar or even unknown, meaning that it could be hard to pinpoint exactly which app was being used. Although looking at what the participant was doing on their phone with their fingers (eg, typing, swiping, tapping) could help distinguish between certain apps that otherwise look quite similar, an extra source of information could be called upon: the participant's own phone. Such information can always be called upon in the moment itself and does not require preprocessing.

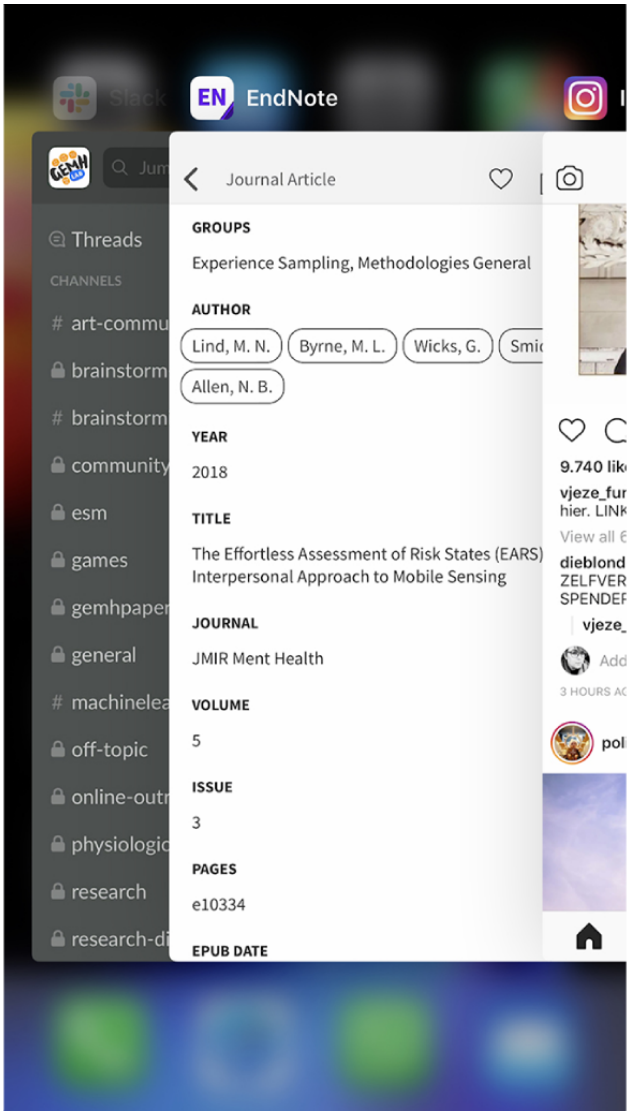
First, if there was uncertainty about what sort of action a participant engaged in, they were asked to open up the social media app and navigate to the activity log or equivalent. Most social media apps contain such an overview of user behavior in the app, although not all of them will refer to this overview as an activity log, and in some cases, information may be scattered over a number of places within the app. For instance, Instagram has an overview of the posts a user has liked, if you dig deep enough, but Instagram does not offer an in-app overview of any comments the user may have posted (one could, however, use the less instantaneous Download My Data functionality if the comments were of particular interest). Luckily, the act of commenting could easily be identified on the video recording since the participant was typing. Facebook, on the other hand, does include comments in their overview of the user's activity, although their activity log is similarly hard to find for an inexperienced user. See Figure 2 for a screenshot of Facebook's activity log. This overview can be helpful for determining which data one can and cannot access in case a similar paradigm is implemented in other studies. Knowing beforehand what sorts of reliable (ie, objective) data can be accessed is paramount for study success, since the stimulated recall hinges on data to aid the participant in accurately recalling thoughts and feelings about the activities of interest.

Figure 2. Screenshot from Facebook app illustrating the activity log.



Second, if there was uncertainty about what sort of app the participant was using, we attempted to retrieve this information using the overview of currently opened apps (see Figure 3 for a screenshot of what that looks like on an iPhone). To make sure that in such a situation we would not be faced with a sequence of apps that had been used days ago (but never closed) rather than in the 10 minutes of the waiting period, we followed a standardized procedure. During the setup of our physiological equipment used to measure participants' electrocardiography (which we incorporated in the study to be able to check whether the stress manipulation had indeed worked), we told the participants that we would need them to turn their phones off and on again so we could do "signal calibration checks" in between to ensure that the phone's signal would not hamper the physiological data collection later on in the study. This meant that we could check, if necessary, which apps in the opened apps overview of the participant's phone were opened and used during the monitoring period and not before; although the app overview remains unchanged even after restarting the phone, any apps being used before the restart will need to reload when accessed from within this overview. That way, we had a way of checking whether an app was used during our study or before it. The "turning phone off and on again" request additionally meant that we could subtly check whether participants had their phone with them and ensure their phone would be near them when the waiting period arrived.

Figure 3. Screenshot of the opened apps overview on an iPhone.



Stimulated Recall Interview

Whereas the use of objective data sources addresses the lack of reliable information regarding user activities, the stimulated recall interview tackles the lack of attention to the how and why of adolescent social media use while incorporating the collected objective data. There are a number of important elements to the successful application of such an interview in a social media research context. First, it is important to acknowledge there are users of social media (especially the younger user base) who engage with social media by sharing relatively personal details about their daily life with friends or family. In order for an interview to be successful, trust needs to be established between researcher and participant because information discussed in the interview, namely about what is put out on social media, can be personal and sensitive. We propose that a powerful way to establish this trust is to authentically recruit the participants' own intrinsic curiosity and generosity in the interview by asking

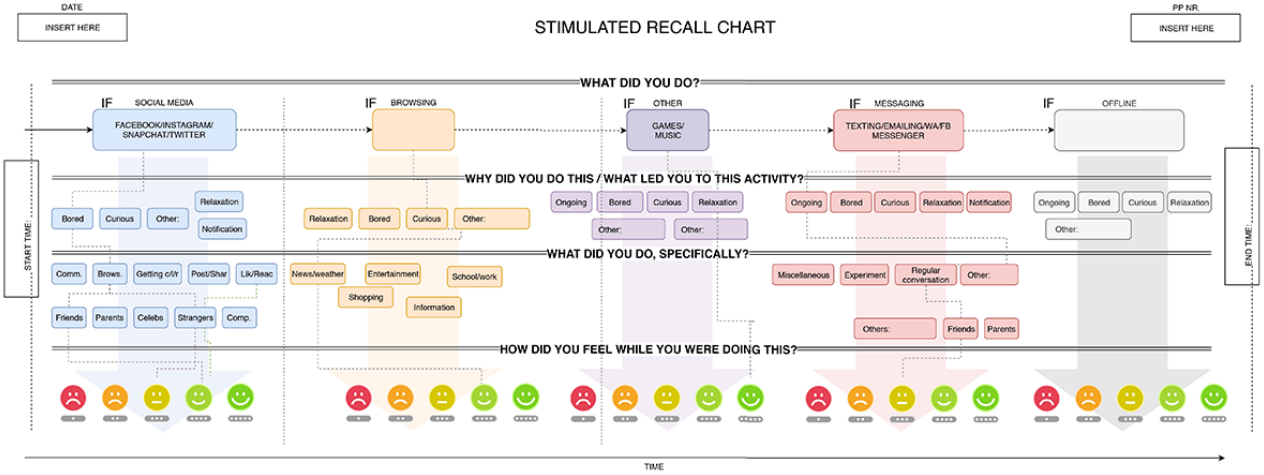
them to join the researcher in a brief moment of co-research. As in participatory research [50], we clearly explain the general goals of our study, what kinds of data have been gathered, and how they will be used to aid in the interview. This is an important step toward eliminating any unease the participant may experience when asked to share personal details, thoughts, and feelings. Second, the interview needs to be structured and standardized across participants. Explaining the structure of the interview will not only help put the participant at ease if necessary, it will also enable the participant to be the best co-researcher they can be; if they know what the researcher is interested in, they will be best able to help and contribute. Good structure and standardization of the interview, however, does not only have to do with the fact that such interviews can be very data-rich. Thanks to a structured approach, the participant will feel there is a particular method and consistency to how personal details are being collected and handled, which will further contribute to a good relationship during the interview. Ultimately, a better researcher-participant relationship will lead to better insights into participants' behaviors and thought processes.

In our study, participants were debriefed and told the true purpose of the study after they had completed the monitoring period in which their activities were recorded. It was explained to participants that the researcher would like to use the remaining study time to conduct a structured and detailed interview about their social media behaviors and experiences, if possible with the aid of the participant's phone and the video recording made during the monitoring period. If the participant gave consent at this point, the study continued and the participant was interviewed following our interview protocol and with aid of the data. If not, the participant was thanked for participation so far and told that the study was ending there. Interestingly, only one of our participants so far has withheld consent for the use of video footage, suggesting participants are interested in sharing their data with us and gaining insight into their own behaviors.

The goal of the interview was to gain insight into (1) what adolescents do on their phones, with increased specificity when it comes to social media, (2) why adolescents engage in these activities (according to them), (3) who (or whose information/posts) they encounter and interact with on social media, and (4) how these activities make adolescents feel in that moment. As such, there were a number of layers to each activity/experience we explored in the interview, and to ensure a consistent structure across participants, we developed a scheme to aid us in conducting these intensive and often personal interviews (Figure 4). By making clear to the participants that despite the personal nature of our questions there was a structure to our method, we hoped to not only facilitate data processing afterward but also predispose participants to cooperate in the interviews (only one participant out of the 53 tested so far has withheld consent for the stimulated recall interview; for more about feasibility research see Feasibility and User Research). Additionally, the interview was audio recorded for future reference and potential in-depth analyses.

During the stimulated recall interview, the video recording of the participant was viewed by the participant and researcher together. At the onset of each new major activity, the researcher would ask the participant questions according to the interview scheme (Figure 4), pausing the video when necessary and completing each of the interview layers (indicated by the horizontal layers in the interview scheme) before moving on to the next major activity. If these behaviors were on the phone, the end/start of a major activity was signified by switching to another app.

Figure 4. Schematic interview chart used to aid in the stimulated recall interview.



The way the interview is structured, going through the layers outlined in the interview scheme, lends the stimulated interview implemented in our study affordances that are especially important for the field of social media (and well-being) research. First, use of video footage of the participants' activities allows us to consider time; research has indicated that we cannot expect people to accurately recall what was done [29], let alone accurately recall in what order. With the video footage as a foundation to construct our image of people's smartphone and social media behaviors, we can pinpoint time stamps to in-app

behaviors (eg, liking a post, reading/scrolling, typing a message, browsing the internet), app switching behaviors, and behaviors like switching from using the smartphone to doing something offline (eg, reading). This allows for the measurement of relatively unexplored variables such as behavior pattern dynamics (eg, whether participants engage in long bouts or short bursts of different activities or whether activities are triggered by incoming notifications or self-initiated), and importantly, allows for accurate assessments of the duration and frequency of behaviors. When viewing the video footage, a first look is taken at the major activity (eg, Facebook). The participant is asked to describe why—to the best of their recollection—they started engaging in this major activity in that moment. Participants may indicate they had a specific goal in mind (eg, "I wanted to look up the profile of a girl a friend mentioned") or they were simply bored and they always go to Facebook when bored. Next, the footage of the major activity is re-viewed and dissected into subactivities done by the participant. For instance, one participant may have scrolled the news feed, liked a number of posts, and commented on one of those posts, whereas another participant might stick to only scrolling the news feed. These subactivities are noted for each major activity. Note that the same major activity may occur multiple times, since people often switch to other apps but then come back afterward. Thanks to the video footage, we are able to capture such repetitions and any differences in behavioural pattern shapes that may occur between participants, giving a much needed, detailed view of how exactly adolescents interact with their phones and social media.

A second advantage to this method is that the interview setting allows us to put the social back into social media research and offers a much more in-depth assessment of whom social media users are engaging and interacting with on these platforms. Social media are, of course, meant to enable social interaction between people all over the world. Additionally, the types of people (eg, family members, friends, acquaintances, strangers, celebrities) users come across on social media may differ vastly depending on the platform. These intricacies of social media use have thus far been ignored in many studies of social media and well-being and can now be addressed in the stimulated recall interview. For each type of subactivity within a major activity (eg, liking posts in a particular Facebook session), we ask participants what type of other people were involved in the subactivity (eg, "What kinds of people posted the messages that you liked?"). The participant is offered a number of suggestions (eg, "Were these messages posted by a friend of yours or by a stranger maybe?") and if possible, the participant's phone is used during the process so that the participant can accurately recover who, for instance, posted the messages they liked. After the participant describes what these people are to them, one of 6 categories is written down next to the subactivity involved: friends, family, romantic partner, acquaintances, strangers, or celebrities.

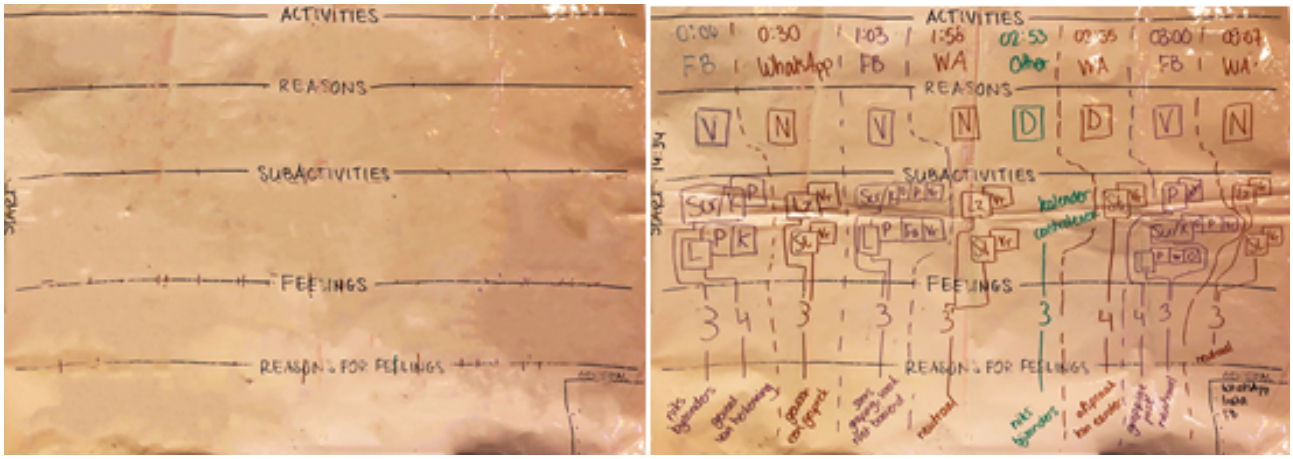
A final advantage to our approach is the in-depth, qualitative nature of the data we can collect. This includes what users did on social media and who they interacted with but also how these behaviors and experiences made them feel. Asking someone to describe the experience of reading a friend's post is hard to do without the context of the post itself, and although there have been studies attempting to artificially recreate such contexts [51], this method provides a more reliable, ecologically convincing account of social media interaction as they emerged spontaneously in a naturalistic context. In our stimulated recall interview, we do not ask participants to elaborate on their feelings for every post they read or emoticon reaction they gave but, given their stimulated recollection of what they read or did, elaborate on their feelings and experiences for the types of activities within the major activity at hand (eg, for scrolling and viewing posts within Facebook). For each of these activities, they indicate with a smiley how they felt in a general sense (on a 5-point Likert scale; see the smileys on the bottom of the interview scheme in Figure 4). After participants indicated which smiley best reflected their feelings for a given activity, we asked participants to briefly describe why they chose this particular smiley and explain how they felt specifically. For instance, if a participant indicated that they felt moderately negative (smiley 2) while scrolling/browsing posts on Facebook, they might say this had to do with the fact that they saw a lot of negative news and it made them a little sad. For this emotional layer of the interview, no categories are used (in contrast to the other layers of the stimulated recall interview). Instead, key words used by the participant when describing how they felt and why are written down. Given that feeling ratings and descriptions are category-based (eg, for scrolling/browsing posts during this particular session of Facebook) rather than per every post they encountered, it may happen that participants report having felt positive emotions for one post and negative emotions for another post. In such cases, participants are asked to select the smiley that most accurately reflects their average feeling about the posts (for instance, by selecting the neutral smiley). The details can always be reflected in the feeling description: it might, for instance, say "It was nice because I saw a funny post, but also sad because a friend of mine had some bad news." While the feeling rating in the form of a smiley is a compact measure, the description element within the feelings layer of the interview plays a vital role in truly finding out how adolescents feel during their social media visits and for which reasons.

Stimulated Recall Chart

The stimulated recall interview yields a rich body of information and data about participant behaviors and experiences. When developing the design of the study and particularly the stimulated recall interview, we developed a standardized chart to formalize all the information coming forth from the participatory interview. After a couple of iterations, we landed on a layout that closely resembles the schematic interview chart (Figure 4). At the start of the interview, a sheet of whiteboard foil with the general skeleton of the chart already set up (Figure 5, left panel) is explained to the participant, and we say we would like the interview to be collaborative and have the participant engage in the research that happens during the interview, together with the researcher. As the interview progresses, we fill out the sheet together, which results in an information-dense but highly structured visualization of the participant's monitoring phase (Figure 5, right panel).

Depending on the activity level of the participant, multiple sheets may be used to capture all of the activities they participated in during the testing phase. After the interview, photos are taken of the sheets to be stored on our secured data servers, and the whiteboard foil sheets are wiped clean, removing all but the initial skeleton of the chart, ready to be used for the next participant.

Figure 5. Left panel: empty whiteboard foil sheet prepared for the stimulated recall interview with the basic skeleton already drawn on it. Right panel: completed example of the stimulated recall chart. Major activities are indicated in the top row; reasons to engage in these activities are indicated in the second row (N: due to a notification; V: boredom); specifics surrounding the activities are indicated in the third row (Scr/K: passive viewing; L: like); and the fourth and fifth rows contain information about participants' feelings during these activities (on a scale from 1-5, with a brief description).



Feasibility and User Research

Monitoring/Waiting Period

Participants were kept unaware of the aim of the study to allow for an optimally naturalistic assessment of adolescents' activities during the waiting period. We expected, from personal and anecdotal experience, that adolescents would pull out their smartphones when asked to wait, and indeed, a pilot we conducted with this method (N=8, all female) indicated that the smartphone was participants' go-to activity. We found an overwhelming display of smartphone use despite the participants' bags being close enough for them to engage in other activities they may have had brought with them (eg, reading a book). This latter aspect of the design (ie, bringing the participant's bag close) was also piloted, since we wanted to give the participant the feeling they could do whatever they wanted (as long as they remained seated) while not diminishing our chances of capturing the behavior of interest (ie, smartphone and social media use). Ultimately, 100% (n=8) of the pilot participants used their phones. Moreover, very few engaged in nonsmartphone activities; only one pilot participant engaged in one offline activity (ie, not involving the phone) in addition to a number of smartphone activities. This participant put away their phone after a couple of minutes and spent the rest of the time investigating the room. Interestingly, this was an older participant (age 54 years), which might explain the difference in behavior compared with our other pilot participants, who were all within the age range of interest (18-25 years). As anticipated, social media were used by nearly all of our pilot participants during the waiting period (again with exception of the participant aged 54 years); in our study, this has been true for all but 9 participants.

Whether the waiting period would allow us to capture the behavior of interest was not the only reason for piloting our paradigm. The duration of the waiting time proved to be a nontrivial issue. Prior to the start of the pilot, durations of 10, 15, and 20 minutes were discussed. We wanted to ensure the participants had enough time to exhibit the full range of possible activities (within the constraints of our study): we didn't know whether participants would go straight to more leisurely activities or attempt to do study-related work first (given that our participants were likely to be students). On the other hand, we did not want the waiting time to affect participants negatively (since waiting too long can be perceived as annoying and might affect the mood). We settled on 15 minutes for the pilot and discovered two issues that directly affected our study. One pilot participant opened an app such as Netflix shortly after the researcher left and continued to watch streaming content for the rest of the waiting period. This, of course, posed a problem for our paradigm, given that we wanted to maximize our chances of capturing social media behaviors. Also, we quickly realized that a 15-minute waiting period (which by definition yielded a 15-minute monitoring video) significantly prolonged the duration of the stimulated recall interview that followed. After a couple of pilot participants, we found that 5 minutes of monitoring footage would take approximately 15 minutes to interview with the participant, bringing the total interview time to 45 minutes in the case of a 15-minute waiting period. These two factors (ie, the predisposition to watch streaming content when told one has to wait for 15 minutes and interview length being a multiple of the waiting period duration) led us to pilot the remaining participants with a monitoring duration of 10 minutes instead. The remaining pilot participants did not engage in streaming series or films, and the interview duration was brought back to 30 minutes, which proved to be more palatable for the participants given the intensive nature of the interview.

Participant Experience

As discussed earlier, the stimulated recall interview could be considered intensive due to the interview's duration and the level of detail of recollection required. To ensure that stimulated

recall paradigms like ours can be used in further studies, experimenters need to make sure the burden on the participants is not too heavy and participants are willing to share their often personal information. Throughout the final phase of our pilot (after the design changes discussed earlier had been consolidated) we asked participants off the record what they thought of the interview and if they would change anything about it, and none of our participants indicated that the interview was too burdensome.

Additionally, although they were often surprised that they had been filmed without their knowledge, all of our pilot participants (n=8) were very willing to share with us how they had spent their time on social media, despite the personal nature of these experiences. We believe that this had a lot to do with the researcher's attitude: transparency is very much valued, even if it is after the fact, and including the participant as much as possible in the process (rather than just using their data) is likely what predisposed our participants to engage in open-hearted conversations during the interview. The emphasis on co-research has also made it easier for our researchers to conduct such personal interviews (rather than feeling like intrusive voyeurs).

With the exception of one participant in our currently running study (n=53) who withheld consent for the use of the monitoring footage, we have found the same as in the pilot: if approached in an open way with genuine interest, participants are excited to explore their own social media activities and experiences with us. For this study, we have decided to include two questions in our debrief measure regarding participant experience in the stimulated recall interview. First, participants were asked to describe in a few sentences how they experienced the stimulated recall interview. Second, participants were asked to indicate on a scale from 1 (very unpleasant) to 10 (very pleasant) what they thought of the stimulated recall interview. Based on the currently collected sample of participants (n=53), we can say that on the whole participants found the method rather pleasant (mean 7.46 [SD 1.27]). The words most used (Figure 6) to describe the stimulated recall method were nice (15), funny (11), interesting (8), great (8), and confrontational (7). These data seem to indicate participants were engaged and enjoyed the process, even though seeing themselves engage in activities that some participants described as not particularly productive was considered confrontational by some.

Figure 6. Word cloud depicting words used by participants (n=53) in response to the statement "Please describe how you experienced the stimulated recall interview." Larger words represent more frequently used words. Font darkness is for decorative purposes only and does not represent other aspects of these data.

Considerations

The application of stimulated recall to social media research is showing promise in helping researchers investigate the functions and processes surrounding social media use by adolescents, while guaranteeing data are reliable. This addresses two important issues currently discussed: the use of self-report for assessment of quantitative aspects of social media use and the lack of attention to the contexts in which social media are used. With the stimulated recall chart as a rich source of data, research questions can be addressed that were out of our reach before. For instance, we will be able to examine students' key motivations for using social media and determine whether use of social media is fueled by the apps themselves (ie, because notifications come in) or by a form of intrinsic motivation. We will also be able to examine in much greater detail what exactly students spend their social media time on and what kinds of social ties (eg, friends, family members or celebrities) they come into contact with on these platforms. Last, and most elusively, we will be able to get a sense of how all of these aspects of

social media use intertwine and culminate in user experiences; how social media use makes people feel likely has to do with the specifics of their use, and we will be able to link the affective experience of social media use to specific aspects of social media use thanks to data gathered through the stimulated recall interview and chart. Importantly, we are not suggesting that stimulated recall should be used exclusively in future research into social media use and well-being. This method, in principle, lends itself well to many other fields, whether the study subject concerns other kinds of new media use (eg, video gaming) or consumer behavior. We suggest that, depending on the research questions at hand, the field can benefit from combining methods such as experimental designs with a stimulated recall approach because this approach is largely content-agnostic: any type of objective data can be used as long as it aids participants in their recall of certain events or experiences. Such multimethod paradigms, if constructed in ecologically valid and reliable ways, will be important steps forward in the field of social media use and well-being.

Despite the clear affordances of this new method, limitations and considerations should be addressed by anyone implementing a similar paradigm in the future. First, the method we have presented is restrictive in that it relies on a setup with cameras. As we found during the piloting phase, placement of the cameras is crucial to the stimulated recall interview because objective sources of information are required to aid the participant in their recall of events. While one could think of nonlaboratory situations (eg, malls, cafes) in which multiple cameras could either be installed or are already present, such contexts bring with them other problems. Privacy issues would arise concerning all other people present in that space who are also captured by the cameras. Recording footage of people without prior consent could be considered ethically unacceptable by institutional review boards. Hence, the stimulated recall method as we have described it here is more suited to controlled settings than field contexts, although the latter might be possible if additional measures are taken to protect the privacy of everyone involved.

Second, stimulated recall interviews do not capture dynamics of behavior over longer periods of time. The monitoring duration is restricted by the following length of the stimulated recall interview and the fact that participants need to wait in one specific lab room. This means only a relatively short period can be captured. To somewhat alleviate this limitation, researchers could adjust the level of detail addressed in the stimulated recall interview (and the data following from that interview); if fewer aspects of behaviors are of interest (eg, only whether something was posted or shared on social media), the stimulated recall interview can be cut short significantly. The stimulated recall interview is a relatively time-intensive form of measurement and, although the timeframe assessed may stretch to 30 minutes or even an hour if that level of detail is needed, the method is simply not suited to assess behaviors (and changes in behaviors) on a time scale that spans multiple days or even weeks.

Third, we have not asked participants to use their phones during the waiting period, and their use of phone and social media can therefore be considered quite natural. However, it should be noted that in our setup, participants are not offered alternative choices of activities, which makes the situation different from usual private life, in which one may be able to choose from using their phone, or reading a book, or watching television. We feel that our setup sufficiently resembles and represents many spare moments in everyday life when roaming public spaces, such as waiting for a friend at a café or riding public transportation.

Fourth, although we have tried to incorporate the best objective data sources available to us in social media contexts (ie, external recordings of the phone and in-phone information), there is still a great amount of data stored within apps and on company servers that remain inaccessible to researchers, mostly due to restrictive data policies asserted by large tech companies. Such data, were they accessible, would be able to shed light on more extended versions of questions that are now assessed using this paradigm. For instance, patterns of behaviors could be assessed over longer periods of time because they could be passively sensed rather than recorded actively in the lab. Although the qualitative, experiential aspect of such behaviors cannot optimally be addressed remotely and will still require an interview with a researcher, reliably measured changes in objective aspects of digital technology use behaviors could be addressed. Additionally, changes in mood could be measured through a complementary experience sampling setup [52], which could then be linked to specific and reliable behavioral data. Efforts toward passive sensing of smartphone use are actively being made [53], but hurdles remain: the fact that many companies do not make their data available to researchers continues to hinder researchers in reliable assessment of users' behaviors. Related to this inaccessibility of data is the fact that users' privacy needs to be ensured no matter what, and, fortunately, efforts toward transparent and privacy-safeguarding protocols are already being made.

Last, we would like to address some ethical considerations pertaining to this kind of protocol. What adolescents do on their phones and on social media can, of course, be highly sensitive. Recording people without their knowledge and consent, especially as they engage in activities generally considered private, is not ideal from an ethical point of view. However, we feel such a lie by omission is necessary to ensure participants display naturalistic behaviors. This type of design is similar to any requiring some form of deception that is later revealed to research participants. However, we take these ethical considerations seriously and have taken steps to minimize concerns and ensure participants provide informed consent as soon as possible during our procedure. Participants are asked for explicit consent a second time, between the collection and use of the data; they are clearly told that they should feel free to withhold consent; they are also assured that they will receive compensation for their time in our lab no matter their decision.

Finally, it is worth mentioning that the nature and extent of precautions taken with regard to storage may depend on the characteristics of the data collected. As we mentioned earlier, participants' faces are not shown in our video footage, and no text or people can be discerned on their phones. However, if the exact content of apps and messages can be read and contacts' faces are recognizable, these factors bring with them new challenges such as ensuring the privacy of those contacts is being guaranteed or their consent obtained.

Conclusions

More attention should be paid to the qualitative side of digital technology use, since frequency and duration metrics can only tell us so much and that seems not to be enough [54]. We have presented a novel paradigm that can be implemented in digital technology (eg, smartphone and/or social media) use research. The application of stimulated recall to these contexts allows us to not only more reliably assess user behaviors but also to address how users think and feel while interacting with digital technologies. We hope that through this paradigm, new insights into people's digital lives can now be gathered in contexts that are ecologically valid and honor the spontaneous and automatic nature of the behavior of interest. Although there are limitations to the stimulated recall paradigm, depending on the question of interest, these can be justified by the insights this method can provide. With more concrete and detailed information about the ways in which users engage with social media and how this makes them feel, researchers in this field will be able to design better studies in the future (whether they be experiments, diary studies, or observational studies) that are ecologically valid and maintain the context sensitivity necessary to capture or create naturalistic behaviors. Ultimately, we hope this new approach will help researchers work toward a better understanding of why, with whom, and how exactly users interact with digital technologies such as social media and how these experiences affect users' mental health and well-being.

Acknowledgments

This work was supported by a European Research Council Consolidator Grant (683262) to IG.

Conflicts of Interest

None declared.

References

1. Young adults are still the most likely to use social media. Washington: Pew Internet and American Life Project; 2015 Oct 08. URL: https://www.pewinternet.org/chart/social-media-users-over-time-by-age/ [accessed 2019-11-18]
2. Twenge JM, Campbell WK. Media use is linked to lower psychological well-being: evidence from three datasets. Psychiatr Q 2019 Jun;90(2):311-331. [doi: 10.1007/s11126-019-09630-7] [Medline: 30859387]
3. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. Clin Psychol Sci 2017 Nov 14;6(1):3-17. [doi: 10.1177/2167702617723376]
4. Dhir A, Yossatorn Y, Kaur P, Chen S. Online social media fatigue and psychological wellbeing: a study of compulsive use, fear of missing out, fatigue, anxiety and depression. Int J Inform Manag 2018 Jun;40:141-152. [doi: 10.1016/j.ijinfomgt.2018.01.012]
5. Hardy BW, Castonguay J. The moderating role of age in the relationship between social media use and mental well-being: an analysis of the 2016 General Social Survey. Comput Hum Behav 2018 Aug;85:282-290. [doi: 10.1016/j.chb.2018.04.005]
6. Kim HH. The impact of online social networking on adolescent psychological well-being (WB): a population-level analysis of Korean school-aged children. Int J Adolesc Youth 2016 Jun 16;22(3):364-376. [doi: 10.1080/02673843.2016.1197135]
7. Steers MN, Wickham RE, Acitelli LK. Seeing everyone else's highlight reels: how Facebook usage is linked to depressive symptoms. J Soc Clin Psychol 2014 Oct;33(8):701-731. [doi: 10.1521/jscp.2014.33.8.701]
8. Weinstein E. The social media see-saw: positive and negative influences on adolescents' affective well-being. New Media Soc 2018 Feb 21;20(10):3597-3623. [doi: 10.1177/1461444818755634]
9. Worsley JD, Mansfield R, Corcoran R. Attachment anxiety and problematic social media use: the mediating role of well-being. Cyberpsychol Behav Soc Netw 2018 Sep;21(9):563-568. [doi: 10.1089/cyber.2017.0555] [Medline: 30132681]
10. Zeeni N, Doumit R, Abi Kharma J, Sanchez-Ruiz M. Media, technology use, and attitudes: associations with physical and mental well-being in youth with implications for evidence-based practice. Worldviews Evid Based Nurs 2018 Aug;15(4):304-312. [doi: 10.1111/wvn.12298] [Medline: 29763998]
11. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: a systematic narrative review. Child Youth Serv Rev 2014 Jun;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]
12. Erfani SS, Abedin B. Impacts of the use of social network sites on users' psychological well-being: a systematic review. J Assoc Inf Sci Technol 2018 Feb 23;69(7):900-912. [doi: 10.1002/asi.24015]
13. Levenson JC, Shensa A, Sidani JE, Colditz JB, Primack BA. The association between social media use and sleep disturbance among young adults. Prev Med 2016 Apr;85:36-41 [FREE Full text] [doi: 10.1016/j.ypmed.2016.01.001] [Medline: 26791323]
14. Jang K, Park N, Song H. Social comparison on Facebook: its antecedents and psychological outcomes. Comput Hum Behav 2016 Sep;62:147-154. [doi: 10.1016/j.chb.2016.03.082]
15. Rui J, Yu N, Xu Q, Cui X. Getting connected while aging: the effects of WeChat network characteristics on the well-being of chinese mature adults. Chinese J Comm 2019;12:25-43. [doi: 10.1080/17544750.2018.1530270]

16. Hall J, Johnson R, Ross E. Where does the time go? An experimental test of what social media displaces and displaced activities' associations with affective well-being and quality of day. New Media Soc 2019;21:674-692. [Medline: https://doi.org/10.1177%2F1461444818804775]
17. Han X, Han W, Qu J, Li B, Zhu Q. What happens online stays online? Social media dependency, online support behavior and offline effects for LGBT. Comput Hum Behav 2019;93:91-98.
18. Phu B, Gow A. Facebook use and its association with subjective happiness and loneliness. Comput Hum Behav 2019;92:151-159 [FREE Full text]
19. Wood M, Center H, Parenteau S. Social media addiction and psychological adjustment: religiosity and spirituality in the age of social media. Ment Health Relig Cult 2016;19:972-983. [doi: 10.1080/13674676.2017.1300791]
20. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: experimental and longitudinal evidence. J Exp Psychol Gen 2015 Apr;144(2):480-488. [doi: 10.1037/xge0000057] [Medline: 25706656]
21. Chow T, Wan H. Is there any "Facebook depression"? Exploring the moderating roles of neuroticism, Facebook social comparison and envy. Pers Individ Dif 2017;119:277-282. [doi: 10.1016/j.paid.2017.07.032]
22. Faelens L, Hoorelbeke K, Fried E, De Raedt R, Koster E. Negative influences of facebook use through the lens of network analysis. Comput Hum Behav 2019;96:13-22. [doi: 10.1016/j.chb.2019.02.002]
23. Hicks S, Brown A. Higher Facebook use predicts greater body image dissatisfaction during pregnancy: the role of self-comparison. Midwifery 2016 Sep;40:132-140. [doi: 10.1016/j.midw.2016.06.018] [Medline: 27428109]
24. Shaw A, Timpano K, Tran T, Joormann J. Correlates of facebook usage patterns: the relationship between passive Facebook use, social anxiety symptoms, and brooding. Comput Hum Behav 2015;48:575-580. [doi: 10.1016/j.chb.2015.02.003]
25. Tandoc EC, Ferrucci P, Duffy M. Facebook use, envy, and depression among college students: is facebooking depressing? Comput Hum Behav 2015:139-146. [doi: 10.1016/j.chb.2014.10.053]
26. Weinstein E. The social media see-saw: Positive and negative influences on adolescents' affective well-being. New Media & Society 2018 Feb 21;20(10):3597-3623. [doi: 10.1177/1461444818755634]
27. Yang C. Instagram use, loneliness, and social comparison orientation: interact and browse on social media, but don't compare. Cyberpsychol Behav Soc Netw 2016 Dec;19(12):703-708. [doi: 10.1089/cyber.2016.0201] [Medline: 27855266]
28. Boase J, Ling R. Measuring Mobile Phone Use: Self-Report Versus Log Data. J Comput-Mediat Comm 2013 Jun 10;18(4):508-519. [doi: 10.1111/jcc4.12021]
29. Ellis D. Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. Comput Hum Behav 2019;97:60-66. [doi: 10.1016/j.chb.2019.03.006]
30. Ellis DA, Davidson BI, Shaw H, Geyer K. Do smartphone usage scales predict behavior? Int J Hum Comput Stud 2019:86-92. [doi: 10.1016/j.ijhcs.2019.05.004]
31. Apaolaza V, Hartmann P, Medina E, Barrutia J, Echebarria C. The relationship between socializing on the Spanish online networking site tuenti and teenagers' subjective wellbeing: the roles of self-esteem and loneliness. Comput Hum Behav 2013;29:1281-1289. [doi: 10.1016/j.chb.2013.01.002]
32. Chou HG, Edge N. "They are happier and having better lives than I am": the impact of using Facebook on perceptions of others' lives. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):117-121. [doi: 10.1089/cyber.2011.0324] [Medline: 22165917]
33. Davila J, Hershenberg R, Feinstein BA, Gorman K, Bhatia V, Starr LR. Frequency and quality of social networking among young adults: associations with depressive symptoms, rumination, and corumination. Psychol Pop Media Cult 2012 Apr 1;1(2):72-86 [FREE Full text] [doi: 10.1037/a0027512] [Medline: 24490122]
34. Krasnova H, Wenninger H, Widjaja T, Buxmann P. Envy on Facebook: a hidden threat to users' life satisfaction? 2013 Presented at: International Conference on Wirtschaftsinformatik; March 1, 2013; Leipzig.
35. Krasnova H, Widjaja T, Buxmann P, Wenninger H, Benbasat I. Why following friends can hurt you: an exploratory investigation of the effects of envy on social networking sites among college-age users. Inf Syst J 2015;26:585-605. [doi: 10.1287/isre.2015.0588]
36. Lee H, Lee H, Choi J, Kim J, Han H. Social media use, body image, and psychological well-being: a cross-cultural comparison of Korea and the United States. J Health Commun 2014;19:1343-1358. [doi: 10.1080/10810730.2014.904022]
37. Pantic I, Damjanovic A, Todorovic J, Topalovic D, Bojovic-Jovic D, Ristic S, et al. Association between online social networking and depression in high school students: behavioral physiology viewpoint. Psychiatr Danub 2012 Mar;24(1):90-93 [FREE Full text] [Medline: 22447092]
38. Sagioglou C, Greitemeyer T. Facebook's emotional consequences: why Facebook causes a decrease in mood and why people still use it. Comput Hum Behav 2014;35:359-363. [doi: 10.1016/j.chb.2014.03.003]
39. Wang SS. "I share, therefore I am": personality traits, life satisfaction, and facebook check-ins. Cyberpsychol Behav Soc Netw 2013 Dec;16(12):870-877. [doi: 10.1089/cyber.2012.0395] [Medline: 23992473]
40. Appel H, Crusius J, Gerlach AL. Social comparison, envy, and depression on Facebook: a study looking at the effects of high comparison standards on depressed individuals. J Soc Clin Psychol 2015 Apr;34(4):277-289. [doi: 10.1521/jscp.2015.34.4.277]
41. Vogel E, Rose J, Okdie B, Eckles K, Franz B. Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. Pers Individ Differ 2015;86:249-256 [FREE Full text]

42. Shakya HB, Christakis NA. Association of Facebook use with compromised well-being: a longitudinal study. Am J Epidemiol 2017 Feb 01;185(3):203-211. [doi: 10.1093/aje/kww189] [Medline: 28093386]
43. Tobin S, Vanman E, Verreynne M, Saeri A. Threats to belonging on Facebook: lurking and ostracism. Social Influence 2014 Mar 07;10(1):31-42. [doi: 10.1080/15534510.2014.893924]
44. Wenninger H, Krasnova H, Buxmann P. Activity matters: investigating the influence of Facebook on life satisfaction of teenage users. 2014 Presented at: Proceedings of the European Conference on Information Systems (ECIS); 2014; Tel-Aviv.
45. Xie W, Karan K. Predicting Facebook addiction and state anxiety without Facebook by gender, trait anxiety, Facebook intensity, and different Facebook activities. J Behav Addict 2019 Mar 01;8(1):79-87. [doi: 10.1556/2006.8.2019.09] [Medline: 30880400]
46. Yuen EK, Koterba EA, Stasio MJ, Patrick RB, Gangi C, Ash P, et al. The effects of Facebook on mood in emerging adults. Psychol Pop Media Cult 2019 Jul;8(3):198-206. [doi: 10.1037/ppm0000178]
47. Calderhead J. Stimulated recall: a method for research on teaching. Br J Educ Psychol 1981;51:211-217. [doi: 10.1111/j.2044-8279.1981.tb02474.x]
48. Bloom B. Thought-processes in lectures and discussions. J Gen Educ 1953;7(3):160-169.
49. Westenberg P, Bokhorst C, Miers A, Sumter S, Kallen V, van Pelt J, et al. A prepared speech in front of a pre-recorded audience: subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. Biol Psychol 2009 Oct;82(2):116-124. [doi: 10.1016/j.biopsycho.2009.06.005] [Medline: 19576261]
50. Boydell K, Gladstone B, Volpe T, Allemang B, Stasiulis E. The production and dissemination of knowledge: a scoping review of arts-based health research. Forum Qual Soc Res 2012;13(1):1. [doi: 10.17169/fqs-13.1.1711]
51. Lin R, Utz S. The emotional responses of browsing Facebook: happiness, envy, and the role of tie strength. Comput Human Behav 2015 Nov;52:29-38. [doi: 10.1016/j.chb.2015.04.064] [Medline: 26877584]
52. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [Medline: 3655778]
53. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The effortless assessment of risk states (EARS) tool: an interpersonal approach to mobile sensing. JMIR Ment Health 2018 Aug 28;5(3):e10334 [FREE Full text] [doi: 10.2196/10334] [Medline: 30154072]
54. Orben A, Dienlin T, Przybylski AK. Social media's enduring effect on adolescent life satisfaction. Proc Natl Acad Sci USA 2019 May 21;116(21):10226-10228 [FREE Full text] [doi: 10.1073/pnas.1902058116] [Medline: 31061122]

Edited by G Eysenbach; submitted 17.07.19; peer-reviewed by A Orben, M George, M Ferrari; comments to author 12.09.19; revised version received 01.10.19; accepted 20.10.19; published 28.01.20.

Please cite as:
Griffioen N, Van Rooij MMJW, Lichtwarck-Aschoff A, Granic I
A Stimulated Recall Method for the Improved Assessment of Quantity and Quality of Social Media Use
J Med Internet Res 2020;22(1):e15529
URL: https://www.jmir.org/2020/1/e15529
doi:10.2196/15529
PMID:32012075

Review

Usability Evaluations of Mobile Mental Health Technologies: Systematic Review

Yavuz Inal[1*], PhD; Jo Dugstad Wake[2*], PhD; Frode Guribye[1*], PhD; Tine Nordgreen[3], PhD

[1]Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

[2]NORCE Norwegian Research Centre, Bergen, Norway

[3]Psychiatric Division, Haukeland University Hospital, Bergen, Norway

[*]these authors contributed equally

Corresponding Author:
Yavuz Inal, PhD
Department of Information Science and Media Studies
University of Bergen
Fosswinckels Gate 6
Bergen, 5020
Norway
Phone: 47 41301855
Email: yavuz.inal@uib.no

Abstract

Background: Many mobile health (mHealth) apps for mental health have been made available in recent years. Although there is reason to be optimistic about their effect on improving health and increasing access to care, there is a call for more knowledge concerning how mHealth apps are used in practice.

Objective: This study aimed to review the literature on how usability is being addressed and measured in mHealth interventions for mental health problems.

Methods: We conducted a systematic literature review through a search for peer-reviewed studies published between 2001 and 2018 in the following electronic databases: EMBASE, CINAHL, PsycINFO, PubMed, and Web of Science. Two reviewers independently assessed all abstracts against the inclusion and exclusion criteria, following the Preferred Reporting Items for Systematic Review and Meta-Analysis guidelines.

Results: A total of 299 studies were initially identified based on the inclusion keywords. Following a review of the title, abstract, and full text, 42 studies were found that fulfilled the criteria, most of which evaluated usability with patients (n=29) and health care providers (n=11) as opposed to healthy users (n=8) and were directed at a wide variety of mental health problems (n=24). Half of the studies set out to evaluate usability (n=21), and the remainder focused on feasibility (n=10) or acceptability (n=10). Regarding the maturity of the evaluated systems, most were either prototypes or previously tested versions of the technology, and the studies included few accounts of sketching and participatory design processes. The most common reason referred to for developing mobile mental health apps was the availability of mobile devices to users, their popularity, and how people in general became accustomed to using them for various purposes.

Conclusions: This study provides a detailed account of how evidence of usability of mHealth apps is gathered in the form of usability evaluations from the perspective of computer science and human-computer interaction, including how users feature in the evaluation, how the study objectives and outcomes are stated, which research methods and techniques are used, and what the notion of mobility features is for mHealth apps. Most studies described their methods as trials, gathered data from a small sample size, and carried out a summative evaluation using a single questionnaire, which indicates that usability evaluation was not the main focus. As many studies described using an adapted version of a standard usability questionnaire, there may be a need for developing a standardized mHealth usability questionnaire.

(J Med Internet Res 2020;22(1):e15337) doi:10.2196/15337

KEYWORDS

systematic review; mobile; mHealth; mental health; usability evaluation

Introduction

Background

Digital technology for screening, treatment, and management of mental health issues has proliferated in recent years, and a substantial share of these applications is implemented on mobile devices [1]. Firth and Torous [2] argue that mobile technologies are particularly suitable to provide services for behavioral health, such as psychiatry, because of the opportunities for capturing patient behavior, for example, through ecological momentary assessment [3] and providing real-time support, given the omnipresence of mobile devices. According to a 2015 report on mobile health (mHealth) adoption, there were 165,000 mHealth apps available on Google Play and iTunes Store, a third of which focused on dieting, wellness, and exercise, and about a quarter of these concerned disease treatment. One-third of the disease-specific apps were for mental health [4].

The World Health Organization (WHO) [5] acknowledges the potential in mHealth apps for meeting the challenges in reaching universal health coverage, provided the apps are evidenced. This involves critically scrutinizing their "benefits, harms, acceptability, feasibility, resource use, and equity considerations." There are different approaches to scientifically assessing the effect, utility, and usefulness of mHealth apps. From the perspective of medicine and psychiatry, the acceptable way of measuring the effects on mental health is through randomized controlled trials [3,6]. From the perspectives of computer science and human-computer interaction (HCI), a well-established approach to the assessment of technology is to evaluate their usability. As a science, usability is grounded not only in the social and behavioral sciences but also in the science of design [7]; however, poor usability and lack of user-centered design have been described as 2 of the reasons for low engagement with mHealth apps [8], and attrition is considered a generic problem in mHealth [9].

Usability is defined by Nielsen [10] as a "quality attribute that assesses how easy interfaces are to use." Usability evaluation has the purpose of gaining understanding of how easy it is to use an interface, and it is an essential part of systems development [11]. There can be different motivations behind usability evaluations, such as establishing evidence that the interface is usable (summative) or informing the redesign and improvement of the interface (formative). Systems with poor usability can lead to situations of low goal-achievement efficiency or the technology not being used or being rejected. Usability evaluation methods are divided into inspection or heuristic methods and methods that are based on input from user representatives. Usability evaluations are usually undertaken in relation to an interaction design process. In HCI, there is an ideal that the results from the evaluation are used to inform the redesign of the evaluated interface but according to Nørgaard and Hornbæk [12], this is often not the case as, rather surprisingly, the evaluation and redesign often occur independently of each other.

Recently, the scope of usability evaluation has shifted from usability engineering to the more encompassing task of evaluating user experience, including user emotions, values, and motivations [13]. At the same time, digital technology is increasingly being directed at the private and public spheres of the users [14], spreading from the "workplace to our homes, everyday lives and culture" [15]. The real-life contexts in which mobile systems are commonly used are often messy and variable, often involving a social context in which other people are present and different kinds of situations with various physical surroundings, such as on the bus or at home [16]. This variation and unpredictability of the context can constitute a challenge when designing and evaluating mobile systems, in addition to the methodological challenges of conducting trials in-the-wild [17], which are particularly applicable to mHealth technologies. Given the prevalence of mobile phones and the confidence the owners have in using them, it seems that usability evaluations of mobile technology are ideal to carry out as field trials, and there is a call for in-the-wild research studies of the use of mobile technology [15]. Yet, there are obstacles to conduct usability evaluations as field trials, for example, the potential difficulties in recreating the intended use situation, combining traditional usability evaluation techniques such as observation and walk-through, and controlling and accounting for all the variables in the environment [16].

As mentioned above, usability evaluation is tightly connected with interaction design. The affordances of mobile devices pose challenges that are particular to designing mobile apps. Compared with desktop computers, mobile devices have several limitations, specifically related to the mobile context in which they are used, such as connectivity, small screen size, different display resolutions, limited processing capability, power, and methods of data entry [18]. On the contrary, mobile devices offer new interaction modalities, such as gestures and movement, location, scan-and-tilt [19], point-of-view and head tracking [20], multitouch and video projection [21], context and proximity sensing, auditory input, and combinations of these features [22]. The proliferation of mobile devices among people all over the world and the many opportunities they provide for creating novel interaction forms raise the question of whether these features are being used in the design of mHealth apps or whether mobile platforms are mainly considered as a convenient way of delivering information.

Objectives

Given the background described above, the goal of our systematic review was to increase the understanding of how usability is being addressed and measured in mobile interventions focusing on mental health problems, where the interventions are made available using mobile devices. We also examined how participants were recruited and which user representatives were involved in the 42 studies from the literature. The following research questions guided the review:

1. What is the approach to users taken in the studies?
2. What are the objectives and outcomes of the studies?
3. What are the characteristics of the mobile apps in the interventions?
4. Why are apps being developed for mobile platforms?
5. Which research methods and techniques are being used to conduct usability evaluation in the studies?

Methods

Study Design

The scope of our review includes how designers approach usability evaluation of new tools and those that are still being developed, that is, their overall evaluation strategies or research approaches to usability evaluation and the concrete methods and usability scales that are being used. We examined how participants are recruited into studies that conduct evaluations of apps, for example, whether the participants are patients diagnosed with a mental illness or mental health professionals and which user representatives are involved. As an aspect of the research approach, usability can be incorporated in the design process and in assessing a developed tool, which can be in the form of co-design or user-centered design, in which future users take part and influence the design process. We looked at the maturity levels of the systems included in the review and the stated purpose of the evaluation. We also reviewed the articles for which mental health issue systems are being designed. Finally, this review assesses the different approaches to mobility as presented in the selected research articles. The reasons for deploying a mental health intervention on mobile devices vary, but in this study, we are primarily interested in why it is a popular platform for deploying mental health interventions.

Information Sources and Search Strategies

A systematic search covering the scientific literature was performed in the medical databases EMBASE, CINAHL, PsycINFO, and PubMed and the wide-ranging scientific database Web of Science. The search was limited to papers published between January 2001 and the end of December 2018. The results were compared and consolidated after each step. The databases were chosen to ensure that all relevant articles could be included in the review study. Search terms were based on a combination of the following keywords: usability, evaluation, assessment, measure, test, testing, heuristics, mental health, mental illness, mental disorder, psychiatric illness, mobile health, M-health, e-health, internet, cCBT (computerized cognitive behavioral therapy), and computerized CBT. The keywords were combined using the Boolean operators OR and AND. The search was customized for each selected database in accordance with their filtering specifications.

Database Searching Process

This review focused on the 4 areas of usability, evaluation methods, mental health, and mobile digital interventions, and accessed relevant articles throughout the 4 steps. Therefore, the search keywords used included words related to these areas. The first keyword was “usability,” which is the main focus of this review. Second, the keywords related to usability evaluation methods were added to identify the applied assessment methods of articles (eg, evaluation OR assessment OR measure OR test OR testing OR heuristics). Third, the results were refined to include the keywords related to the mental health domain (eg, mental health OR mental illness OR mental disorder OR psychiatric illness). Fourth, search keywords addressing mobile digital intervention in mental health were added to limit the search results and access more relevant articles.

Selection of Studies

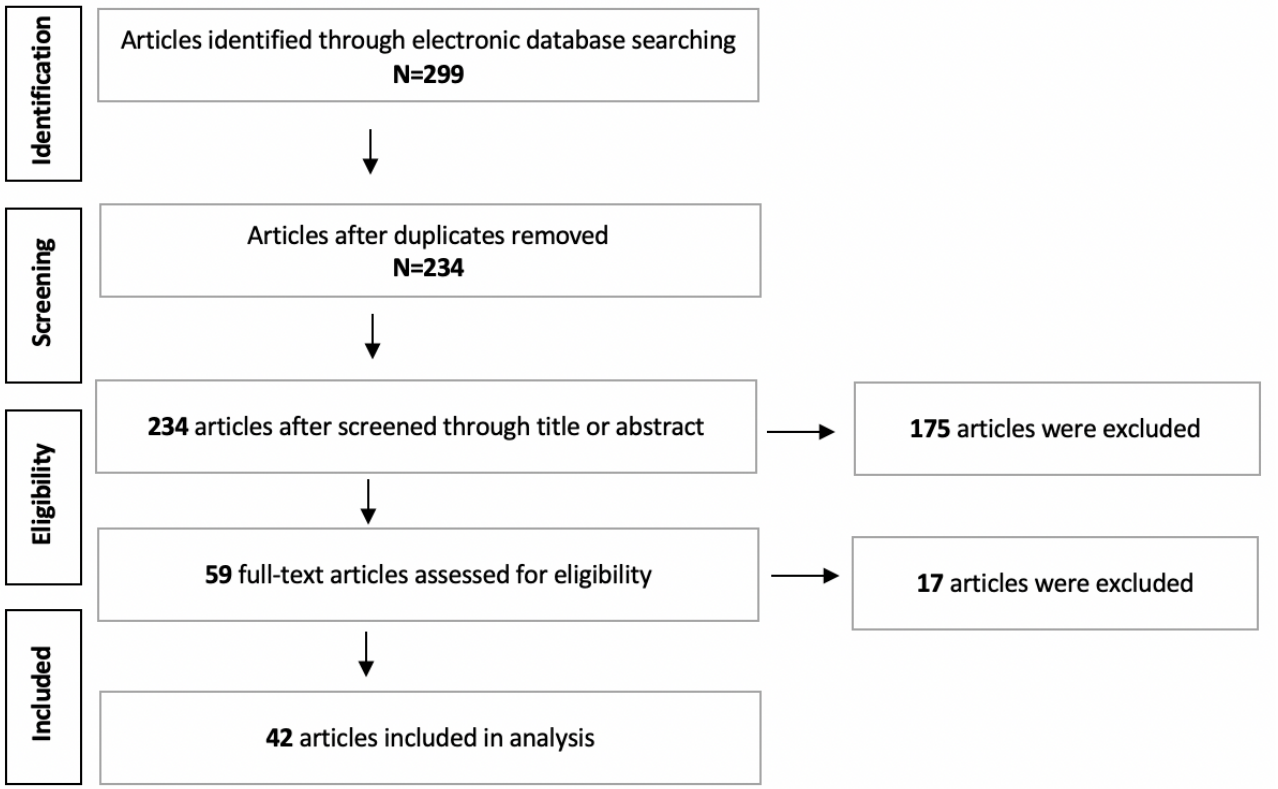
The Preferred Reporting Items for Systematic Review and Meta-Analysis statement was used for the reporting of the systematic review [23]. A total of 5 databases were searched systematically using predetermined keywords. The reference lists of the included articles were also searched for additional relevant articles. After removing duplicates, a set of inclusion and exclusion criteria were formulated to evaluate and identify the most relevant articles.

We included those studies in which the articles met the following inclusion criteria (1) focusing on usability evaluation of a mobile digital mental health intervention and (2) providing empirical evidence with regard to the usability evaluation outcomes of digital mental health interventions. We also excluded the studies that met at least one of these exclusion criteria: (1) not written in English; (2) published before 2001 or after December 2018; (3) not having a full text or published in the form of a conference paper or an abstract; (4) designed as nonempirical research (eg, opinion papers, reviews, editorials, and letters); (5) study protocol; (6) dealing with usability evaluation in domains that do not include mobile digital mental health; and (7) having limited mobile use to SMS, as a Web browsing platform, or purely as a sensor.

The database searches were performed by 2 of the authors independently in a double-blind process. After identifying relevant articles through the electronic database search, 65 duplicate articles were removed, and 234 unique articles remained. In the screening step, the resulting list of 234 articles were reviewed independently by the same 2 authors according to the inclusion and exclusion criteria by considering the title, keywords, and abstract, and all 59 eligible studies were retrieved. To assess the eligibility of the remaining articles, the full texts were evaluated, provided the information given in the abstract was sufficient to decide on the relevance of the article.

The full texts of all identified articles were assessed independently by the same authors. Articles upon which both authors agreed were included. Any discrepancies between the authors regarding the selection of the articles were discussed, and a consensus was reached on all reviewed articles in a joint session. In total, 17 articles were excluded in this round, and the selection process led to the inclusion of 42 articles in this review as shown in Figure 1. The main method to resolve discrepancies was to review the full text paper with regards to IC2: whether the paper described a usability study including empirical evidence.

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow diagram of article inclusion.



Defining the Evaluation Criteria Used in the Study

In accordance with the research questions, the evaluation criteria were grouped according to a set of 4 themes: (1) approach to *the user*, (2) *objectives and outcomes* of the study, (3) *research methods* and techniques to conduct usability evaluation, and (4) information about the *mHealth interventions*. A summary of the themes, evaluation criteria, and their values are given in Table 1.

The theme *approach to the users* contains information that is descriptive of the participants. The mental health problem addressed in the study refers to the stated diagnosis or mental health symptoms that the intervention is directed at. These can be diagnosis-based, symptom-based, and some interventions are also more general in nature, or about *wellness* or providing *access to information*. The *sample size* refers to how many participants took part in the evaluation, and we included a description of their role, that is, being patients, experts, health professionals, relatives, and so on. In *target demographics*, we reviewed whether there were any particular social strata that were addressed through the intervention in addition to the diagnosis, such as gender, age group, and culture.

Table 1. Themes, evaluation criteria, and main subvalues and categories used in the study.

Theme	Evaluation criteria	Main subvalues and categories
Approach to the users	Type of mental health problem/diagnosis, sample size, target demographics	—[a]
Study objectives and outcomes	Purpose of the study, outcome of the study	Outcome: User reception, tool improvement, design recommendations, design themes, value in exploration, medical outcomes (positive, negative, and neutral), research improvement and app/tool
Methods and techniques	Research methods, usability evaluation techniques, and purpose of the usability evaluation	Research methods: Trial, user-centered design, mixed methods study, and participatory design; Evaluation methods: Interview (type), think-aloud, questionnaire (type), field study (natural environment), app use data, co-operative design, verbal probing, observation, scenario-based tasks, focus-group, panel review, video recording, logging/diaries, task-based evaluation, wireframing/sketching, personas; Purpose: Formative, summative
Mental health intervention	Maturity level of the mobile system and approach to mobility	Approach category: Device affordances, availability of mobile technology, contextual support, novelty of mobile research, popularity, user maturity, and privacy of mobile use

[a]The theme has no subvalue or category.

The theme *objectives and outcomes* describe the author's stated purpose of the article. The meaning of purpose was considered self-explanatory. We found statements of objective in the abstract section in most of the papers, where the articles had a subheading in the abstract where this was described explicitly. We also reviewed the introduction and methods sections, in which the objectives of the studies were explained. Outcome descriptions were grouped and categorized according to their meaning. The outcomes refer to conclusions about the main contributions of the study, as described by the authors. The outcomes were categorized using the terms user reception, tool improvement, design themes, value in exploration, medical outcomes (positive, negative, and neutral), research improvement, and app/tool. User reception refers to conclusions about how the technology was received and perceived by the participants in the study. Tool improvement means the conclusions about how the technology was improved are based on the feedback from the users. Design themes are observations that are sufficiently general to be of value to other designers and developers of mobile mental health technologies. Value in exploration means that the authors found that taking an explorative approach to evaluation gave significant knowledge in return for the study. Medical outcomes are conclusions concerning the medical effects on the participants of the study. Research improvement are findings that increase quality of future mHealth usability evaluations. Finally, app/tool refers to conclusions about how the technology was accomplished, usually based on user feedback.

Methods and techniques describe the methodological aspects of the studies. Research methods refer to the overall research strategy employed in the study that the article describes. We decided on the following main types of research methods: trial, user-centered design, mixed methods study, and participatory design. A trial is about determining the effects of an intervention. A criterion for a trial was that the participants used the technology independently for a period either in their natural environment or elsewhere. If the trial took place in the user's natural environment, this was specified in the usability evaluation column. Mixed methods studies refer to a set of qualitative and/or quantitative techniques being used to study the intervention but not in the sense of a trial. User-centered design methods refer to studies where potential users or representatives of users have taken part in the design process, for example, in the form of co-design or participatory design. User-centered design methods also usually entail an iterative design and evaluation process. The purpose of the evaluation describes whether intervention technology redesign based on user input took place in the study, described as formative if so and summative where the focus of the study was on describing the usability evaluation results. Usability evaluation refers to the data collection techniques that were used in the study, that is, interviews, observation, and questionnaires, including standardized usability measures, such as the System Usability Scale (SUS) or their adaptations.

The theme *mHealth intervention* refers to information about digital intervention. By reviewing the *maturity of the technology*, we wanted to see if there were discerning trends in how early in development usability evaluations were carried out. For maturity, we distinguished between the main categories of sketch, prototype, matured, and released version of technology. In *approach to mobility*, we were interested in if and how the authors argued for the choice to use mobile technology for their intervention. The main categories for these were device affordances, availability of mobile technology, contextual support, novelty of mobile research, popularity, user maturity, and privacy of mobile use. Interventions that employ mobile devices because of specific affordances (sensors, communication, etc) use aspects of mobile devices that are difficult or less practical to replicate using other devices. Availability refers to the notion that mobile devices are available to the user most of the time. Popularity means that mobile devices are currently in use by most people for most age groups. User maturity refers to the idea that people have become technologically proficient in mobile phone use. Privacy is related to the notion that people regard, for example, finding information on their mobile phone as more private than visiting a mental health professional.

Results

Approach to the Users

A total of 29 of the studies conducted a usability evaluation of a mobile mental health program with either patients or patient families with parent-child dyads, 11 with health care providers, such as clinicians, caregivers, nurses, therapists, care managers, and health professionals, and 8 with healthy users. Of the studies gathering feedback from users, 1 study recruited users with a history of trauma, encouraged users with lived experience of mental health and substance use to participate, and evaluated usability with users who were offspring of patients with dementia, and another recruited parents of children with neurodevelopmental disabilities. To conduct a more detailed analysis, some of the studies performed the usability evaluation of a mobile mental health program with different user groups. A total of 6 studies obtained usability feedback from both patients and health care providers. The categories healthy users and patients, healthy users and health care providers, healthy users and experts, health care providers and experts, patients and practitioners, patients and teachers of dyslexia, health care providers, patients and researchers on health domain, and health care providers, patients and healthy users occurred once (see Table 2).

Table 2. Types of users recruited by the reviewed studies.

User type	Study
Patients	Auger et al, 2014 [24]; Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; Ben-Zeev et al, 2013 [27]; Ben-Zeev et al, 2014 [28]; Boman and Bartfai, 2015 [29]; Boyd et al, 2017 [30]; Corden et al, 2016 [31]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Henry et al, 2017 [35]; Huguet et al, 2015 [36]; Juengst et al, 2015 [37]; Kobak et al, 2015 [38]; Latif et al, 2015 [39]; Macias et al, 2015 [40]; Meiland et al, 2012 [41]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Nitsch et al, 2016 [45]; Palmier-Claus et al, 2013 [46]; Prada et al, 2017 [47]; Rizvi et al, 2016 [48]; Rohatagi et al, 2016 [49]; Ruggiero et al, 2015 [50]; Sze et al, 2015 [51]; Whiteman et al, 2017 [52]
Health care providers	Bauer et al, 2018 [26]; Boman and Bartfai, 2015 [29]; Fuller-Tyszkiewicz et al, 2018 [34]; Kobak et al, 2015 [38]; Meiland et al, 2012 [41]; Ospina-Pinillos et al, 2018 [53]; Rohatagi et al, 2016 [49]; Ruggiero et al, 2015 [50]; Sands et al 2016, [54]; Villalobos et al, 2017 [55]; Wood et al, 2017 [56]
Healthy users	Boyd et al, 2017 [30]; Carey et al, 2016 [57]; Connelly et al, 2016 [58]; de Korte et al, 2018 [59]; Garcia et al, 2017 [60]; Kizakevich et al, 2018 [61]; Ospina-Pinillos et al, 2018 [53]; Rohatagi et al, 2016 [49]
Users with a mental health history	Price et al, 2016 [62]
Users with lived experience of mental health and substance use	VanHeerwaarden et al, 2018 [63]
Users who were offspring of patients with a mental health illness	van Osch et al, 2015 [64]
Users who were parents of children with a mental health	Jiam et al, 2017 [65]
Researchers on health domain	Fuller-Tyszkiewicz et al, 2018 [34]
Teachers of dyslexia	Latif et al, 2015 [39]
Practitioners	Ben-Zeev et al, 2013 [27]
Experts	de Korte et al, 2018 [59]; Sands et al, 2016 [54]

The total sample size at baseline (regardless of the number of groups) ranged from 5 [24,43] to 3977 [60]. A total of 3 studies reported targeting only females [45,47,58], whereas 1 study gathered data only from male patients [29] and male users [61]. There was an equal gender distribution in 4 studies [25,40,52,55]. One study recruited the same number of males and females in stage 1, but all males in stage 2 [32], and another study included 1 group of users (young people), and not the other group (youth health professional) [53]. Gender information was not reported in 3 studies about health care providers [29,49,50], in 2 about users [43,61], in 1 about teachers of dyslexia [39], in 1 about practitioners [27], in 1 about experts and health care providers [54], and in another about users and health care providers [26]. A total of 8 studies reported only the age range of the participants [36,40,44,45,55,57,63,65], 6 provided only the mean age [25,28,31,34,46,62], and 2 did not provide this information [26,60]. Although some of the studies gathered usability feedback from different user groups, such as both patients and health care providers or patients and healthy users, a considerable number of the studies (n=12) did not present the same level of detailed information about all participants' demographics for each user group, such as the mean age, age range, and gender [27,29,30,35,49-51,53,58,61].

A significant number of the included studies addressed generic mental health issues, such as well-being, mindfulness, and goal achievement, followed by depression, schizophrenia, alcohol use disorder, bipolar disorder, cognitive impairment, eating disorder and serious mental illness, borderline personality disorder, dementia, medical adherence, and posttraumatic stress disorder. The full list of mental health problems is presented in Table 3.

Table 3. Mental health problems addressed in the studies.

Mental health problem	Study
A history of violence	Mistler et al, 2017 [42]
Alcohol dependence and misuse	Barrio et al, 2017 [25]; Kizakevich et al, 2018 [61]; Dulin et al, 2014 [33]
Anger	Morland et al, 2016 [43]
Bipolar disorder	Bauer et al, 2018 [26]; Macias et al, 2015 [40]; Mistler et al, 2017 [42]
Borderline personality disorder	Prada et al, 2017 [47]; Rizvi et al, 2016 [48]
Burnout	Wood et al, 2017 [56]
Cognitive impairment	Boman and Bartfai, 2015 [29]; Boyd et al, 2017 [30]; Auger et al, 2014 [24]
Dementia	Meiland et al, 2012 [41]; van Osch et al, 2015 [64]
Depression	Corden et al, 2016 [31]; Deady et al, 2018 [32]; Fuller-Tyszkiewicz et al, 2018 [34]; Kobak et al, 2015 [38]; Macias et al, 2015 [40]
Eating disorders	Connelly et al, 2016 [58]; Nitsch et al, 2016 [45]; Sze et al, 2015 [51]
Dyslexia	Latif et al, 2015 [39]
Generic (communication access, assessment, contentment, well-being, goal achievement, and mindfulness)	Carey et al, 2016 [57]; de Korte et al, 2018 [59]; Garcia et al, 2017 [60]; Ospina-Pinillos et al, 2018 [53]; Ruggiero et al, 2015 [50]; Sands et al, 2016 [54]; VanHeerwaarden et al, 2018 [63]; Villalobos et al, 2017 [55]
Headache	Huguet et al, 2015 [36]
Medication adherence	Corden et al, 2016 [31]; Rohatagi et al, 2016 [49]
Neurodevelopmental disabilities	Jiam et al, 2017 [65]
Posttraumatic stress disorder	Bauer et al, 2018 [26]; Price et al, 2016 [62]
Psychosis	Palmier-Claus et al, 2013 [46]
Schizoaffective disorder	Mistler et al, 2017 [42]
Schizophrenia	Ben-Zeev et al, 2013 [27], 2014 [28]; Macias et al, 2015 [40]; Mistler et al, 2017 [42]; Palmier-Claus et al, 2013 [46]
Serious mental illness	Whiteman et al, 2017 [52]; Nicholson et al, 2018 [44]; Rohatagi et al, 2016 [49]
Sleep problems	Kizakevich et al, 2018 [61]
Stress	Kizakevich et al, 2018 [61]
Tinnitus	Henry et al, 2017 [35]
Traumatic brain injury	Juengst et al, 2015 [37]

Objectives and Outcomes

Across the studies, the reported primary purposes differed considerably. Half of the studies emphasized usability evaluation [24,25,29-64], 10 focused on feasibility [28,31,32,36-38,42,44,48,51] and acceptability [28,32,40,42,47,48,51,56,60], and for 5, effectiveness [32,33,38,48,56] was the main objective. Some of the studies had the purpose of concentrating on patients attitudes, such as satisfaction [25,38], perception [46], openness [47], motivation [64], opinions [59], and adherence to the use of a mobile mental health app [49], whereas others addressed mobile apps, for example, system usage [33,44], app optimization [63,64], validity of a mHealth system [37], efficacy [28], usefulness [44], perceived quality [60], content validity [54], significant features in content [61], safety [49], psychometric properties [36], and health assessment quality [61].

Numerous studies described the process of design [53,58], development [27,30,32,35,36,49,50,53,65], and adaptation [55] of a mobile mental health app or platform, whereas a few aimed to demonstrate the value of usability research [45] and benefits of mobile technologies in providing a learning platform [39] or examined how to incorporate mobile technologies to support delivery of a mental health service [26]. Only 2 studies targeted to test an intervention [35] and improve the treatment of depression [31].

The outcomes of almost all of the included studies, except one, were user reception, followed by medical outcome (positive), tool improvement, app/tool, design recommendations, design themes, medical outcome (potential), medical outcome (neutral), and product and implementation issues. Outcomes that occurred once were value in exploration, research improvement, medical outcome (indirectly), design principles, and evaluation knowledge. Details are given in Table 4.

Table 4. Outcomes of the included studies.

Outcome	Study
User reception	Auger et al, 2014 [24]; Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; Ben-Zeev et al, 2013 [27], 2014 [28]; Boman and Bartfai, 2015 [29]; Boyd et al, 2017 [30]; Carey et al, 2016 [57]; Connelly et al, 2016 [58]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Garcia et al, 2017 [60]; Henry et al, 2017 [35]; Huguet et al, 2015 [36]; Jiam et al, 2017 [65]; Juengst et al, 2015 [37]; Kizakevich et al, 2018 [61]; Kobak et al, 2015 [38]; Latif et al, 2015 [39]; Macias et al, 2015 [40]; Meiland et al, 2012 [41]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Nitsch et al, 2016 [45]; Ospina-Pinillos et al, 2018 [53]; Palmier-Claus et al, 2013 [46]; Prada et al, 2017 [47]; Price et al, 2016 [62]; Rizvi et al, 2016 [48]; Rohatagi et al, 2016 [49]; Sands et al, 2016 [54]; Sze et al, 2015 [51]; van Osch et al, 2015 [64]; VanHeerwaarden et al, 2018 [63]; Villalobos et al, 2017 [55]; Whiteman et al, 2017 [52]; Wood et al, 2017 [56]
Medical outcome (positive)	Ben-Zeev et al, 2014 [28]; Carey et al, 2016 [57]; Corden et al, 2016 [31]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Garcia et al, 2017 [60]; Huguet et al, 2015 [36]; Juengst et al, 2015 [37]; Kobak et al, 2015 [38]; Macias et al, 2015 [40]; Mistler et al, 2017 [42]; Prada et al, 2017 [47]; Rizvi et al, 2016 [48]; Sze et al, 2015 [51]; Wood et al, 2017 [56]
Tool improvement	Connelly et al, 2016 [58]; Henry et al, 2017 [35]; Jiam et al, 2017 [65]; Meiland et al, 2012 [41]; Nitsch et al, 2016 [45]; Ruggiero et al, 2015 [50]; Sands et al, 2016 [54]; van Osch et al, 2015 [64]; Whiteman et al, 2017 [52]
App/tool	Ben-Zeev et al, 2013 [27]; Connelly et al, 2016 [58]; Deady et al, 2018 [32]; Henry et al, 2017 [35]; Latif et al, 2015 [39]; Ospina-Pinillos et al, 2018 [53]; Ruggiero et al, 2015 [50]; VanHeerwaarden et al, 2018 [63]
Design recommendations	Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Garcia et al, 2017 [60]; Juengst et al, 2015 [37]; Ospina-Pinillos et al, 2018 [53]; Price et al, 2016 [62]
Design themes	Auger et al, 2014 [24]; Connelly et al, 2016 [58]; Nitsch et al, 2016 [45]
Medical outcome (potential)	Latif et al, 2015 [39]; Ruggiero et al, 2015 [50]; Whiteman et al, 2017 [52]
Medical outcome (neutral)	Kizakevich et al, 2018 [61]; Meiland et al, 2012 [41]
Product	Henry et al, 2017 [35]; Ruggiero et al, 2015 [50]
Implementation issues	Boman and Bartfai, 2015 [29]; Palmier-Claus et al, 2013 [46]
Value in exploration	Villalobos et al, 2017 [55]
Research improvement	Macias et al, 2015 [40]
Medical outcome (indirectly)	Boman and Bartfai, 2015 [29]
Design principles	Bauer et al, 2018 [26]
Evaluation knowledge	de Korte et al, 2018 [59]

Characteristics of Mobile Health Interventions

The maturity level of the mobile systems that were reviewed were placed on a continuum from sketch to final product (Figure 2). A sketch-to-prototype means that the study described the development of the app in the form of co-design and that gleaning feedback and worldviews of the users were the focus of the study. A prototype is the minimally working version of an app with functionality that the user can test. A matured version is an app that has been tested by users and redesigned/amended in some way. A released version refers to the app being downloadable from an app store or elsewhere, and the final version is self-explanatory.

The most common maturity level of technology in the review was divided among 13 studies of matured version of technology, 8 studies of released version technology, 9 prototype, and 6 prototype-to-matured. There was only 1 paper that evaluated a final product, whereas 3 studies described the process from sketch to prototype. For 2 of the studies, no information about the maturity level was available (Table 5). For the categories sketch-to-prototype and prototype-to-matured, the studies were focused on describing a development process, where user feedback was used in a formative redesign of technology.

Figure 2. Mobile health technology maturity scale.

Table 5. Maturity levels of the mobile systems that were reviewed.

Maturity level	Study
Sketch	—[a]
Sketch-to-prototype	Ospina-Pinillos et al, 2018 [53]; Sands et al, 2016 [54]; Whiteman et al, 2017 [52]
Prototype	Ben-Zeev et al, 2013 [27]; Carey et al, 2016 [57]; Deady et al, 2018 [32]; Jiam et al, 2017 [65]; Latif et al, 2015 [39]; Nitsch et al, 2016 [45]; Price et al, 2016 [62]; Rohatagi et al, 2016 [49]; van Osch et al, 2015 [64]
Prototype-to-matured	Bauer et al, 2018 [26]; Connelly et al, 2016 [58]; Henry et al, 2017 [35]; Huguet et al, 2015 [36]; Meiland et al, 2012 [41]; Ruggiero et al, 2015 [50]
Matured	Barrio et al, 2017 [25]; Ben-Zeev et al, 2014 [28]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Garcia et al, 2017 [60]; Juengst et al, 2015 [37]; Macias et al, 2015 [40]; Nicholson et al, 2018 [44]; Palmier-Claus et al, 2013 [46]; Sze et al, 2015 [51]; VanHeerwaarden et al, 2018 [63]
Released version	Auger et al, 2014 [24]; Boyd et al, 2017 [30]; Kizakevich et al, 2018 [61]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Prada et al, 2017 [47]; Rizvi et al, 2016 [48]; Wood et al, 2017 [56]
Final version	Boman and Bartfai, 2015 [29]
No information	Kobak et al, 2015 [38]; Villalobos et al, 2017 [55]

[a]Not applicable.

We reviewed the articles concerning how the authors argued for the use of mobile devices, and which and how mobile device affordances were used to make a tool. A summary of approaches to mobility results is given in Table 6. The availability of mobile devices was the most commonly cited reason to develop mHealth tools, which was found in 21 of the articles. The current popularity of mobile devices was mentioned in 16 of the studies, whereas 14 studies referred to or used affordances that are difficult to replicate on nonmobile devices, such as sensors. For 9 of the articles, this affordance was the potential for facilitating communication. A total of 8 papers referred to the novelty of mobile research, that is, it is worth exploring mHealth because it is relatively new and unexplored territory. A total of 8 articles referred to user maturity, meaning that their intended users were proficient in the use of mobile devices, whereas 5 papers mentioned the potential privacy of mobile use. A total of 4 articles pointed to the successful use of mobile technology in previous mHealth research, whereas 2 focused on how mHealth technologies could give the user control over their mental health problems. For 3 of the papers, no information was available. Overall, 1 paper each referred to how mHealth technology could augment existing practices, how it could increase cost effectiveness, and how it could support scalable solutions.

Table 6. Approaches to mobility.

Mobility approach	Study
Availability	Barrio et al, 2017 [25]; Ben-Zeev et al, 2014 [28]; Carey et al, 2016 [57]; Connelly et al, 2016 [58]; de Korte et al, 2018 [59]; Deady et al, 2018 [32]; Garcia et al, 2017 [60]; Henry et al, 2017 [35]; Huguet et al, 2015 [36]; Jiam et al, 2017 [65]; Juengst et al, 2015 [37]; Latif et al, 2015 [39]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Palmier-Claus et al, 2013 [46]; Prada et al, 2017 [47]; Price et al, 2016 [62]; Rizvi et al, 2016 [48]; Sands et al, 2016 [54]; Whiteman et al, 2017 [52]; Wood et al, 2017 [56]
Popularity	Bauer et al, 2018 [26]; Ben-Zeev et al, 2013 [27]; Ben-Zeev et al, 2014 [28]; de Korte et al, 2018 [59]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Garcia et al, 2017 [60]; Kizakevich et al, 2018 [61]; Kobak et al, 2015 [38]; Mistler et al, 2017 [42]; Nicholson et al, 2018 [44]; Ospina-Pinillos et al, 2018 [53]; Prada et al, 2017 [47]; Price et al, 2016 [62]; Rizvi et al, 2016 [48]; Whiteman et al, 2017 [52]
Device affordances	Auger et al, 2014 [24]; Barrio et al, 2017 [25]; Boman and Bartfai, 2015 [29]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Kizakevich et al, 2018 [61]; Latif et al, 2015 [39]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Palmier-Claus et al, 2013 [46]; Price et al, 2016 [62]; Rohatagi et al, 2016 [49]; Ruggiero et al, 2015 [50]; van Osch et al, 2015 [64]
Communication affordance	Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; Boman and Bartfai, 2015 [29]; Carey et al, 2016 [57]; Jiam et al, 2017 [65]; Kobak et al, 2015 [38]; Nitsch et al, [45]; Price et al, 2016 [62]; van Osch et al, 2015 [64]
Novelty of mobile research	Auger et al, 2014 [24]; Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; de Korte et al, 2018 [59]; Juengst et al, 2015 [37]; Macias et al, 2015 [40]; Sze et al, 2015 [51]; Villalobos et al, 2017 [55]
User maturity	Ben-Zeev et al, 2013 [27]; Fuller-Tyszkiewicz et al, 2018 [34]; Jiam et al, 2017 [65]; Juengst et al, 2015 [37]; Nicholson et al, 2018 [44]; Rohatagi et al, 2016 [49]; Sze et al, 2015 [51]; Whiteman et al, 2017 [52]
Privacy of mobile use	Dulin et al, 2014 [33]; Kizakevich et al, 2018 [61]; Macias et al, 2015 [40]; Nicholson et al, 2018 [44]; Ospina-Pinillos et al, 2018 [53]
Scientific evidence of a positive effect	Deady et al, 2018 [32]; Fuller-Tyszkiewicz et al, 2018 [34]; Morland et al, 2016 [43]; Wood et al, 2017 [56]
Contextual support	Ben-Zeev et al, 2013 [27]; Macias et al, 2015 [40]; Villalobos et al, 2017 [55]
No information	Boyd et al, 2017 [30]; Meiland et al, 2012 [41]; VanHeerwaarden et al, 2018 [63]
User control	de Korte et al, 2018 [59]; Palmier-Claus et al, 2013 [46]
Augment existing practice	Sands et al, 2016 [54]
Cost effectiveness	de Korte et al, 2018 [59]
Scalability	Ruggiero et al, 2015 [50]

Research Methods and Techniques of the Studies

Regarding the purpose of the usability evaluation, 31 of the included studies carried out a summative evaluation [24,25,27,64], whereas 11 undertook a formative evaluation [26,32,35-65]. A total of 3 studies carried out both summative and formative evaluations in separate phases [27,50,58]. A total of 32 studies were described as trials [25-27,30,31,33-35,38-65], whereas 12 used the method of user-centered design [24,26,28,32,35,39,41,50,52,58,64,65]. A total of 4 were mixed methods studies [29,44,62], and 3 were described as participatory design [36,53,66].

The most common data collection technique used was a questionnaire, either self-constructed, standard, or combinations of these. A total of 33 studies used questionnaires. In all, 31 studies were conducted as field studies in the natural environment of the participants, with the technology deployed in the everyday environment of the intended future user or the representatives of these users. A total of 23 studies made use of interviews. Less frequently used methods were observation, think-aloud, and the use of app-use generated data, task-based evaluation, and focus groups in order of frequency. The methods that were only referred to once in all the studies in the review were sensor data, co-operative design with both users and experts, verbal probing, user feedback, video recording, diaries, wireframing/sketching, personas, and journey mapping (see Table 7).

Table 7. Data collection techniques employed in the studies.

Outcome	Study
Questionnaire	Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; Ben-Zeev et al, 2013 [27], 2014 [28]; Boman and Bartfai, 2015 [29]; Connelly et al, 2016 [58]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Garcia et al, 2017 [60]; Huguet et al, 2015 [36]; Jiam et al, 2017 [65]; Juengst et al, 2015 [37]; Kizakevich et al, 2018 [61]; Kobak et al, 2015 [38]; Latif et al, 2015 [39]; Meiland et al, 2012 [41]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Nitsch et al, 2016 [45]; Prada et al, 2017 [47]; Price et al, 2016 [62]; Rizvi et al, 2016 [48]; Rohatagi et al, 2016 [49]; Sze et al, 2015 [51]; van Osch et al, 2015 [64]; VanHeerwaarden et al, 2018 [63]; Villalobos et al, 2017 [55]; Whiteman et al, 2017 [52]; Wood et al, 2017 [56]
Field study	Auger et al, 2014 [24]; Barrio et al, 2017 [25]; Bauer et al, 2018 [26]; Ben-Zeev et al, 2013 [27]; Boman and Bartfai, 2015 [29]; Boyd et al, 2017 [30]; Carey et al, 2016 [57]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Deady et al, 2018 [32]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Garcia et al, 2017 [60]; Henry et al, 2017 [35]; Jiam et al, 2017 [65]; Juengst et al, 2015 [37]; Kizakevich et al, 2018 [61]; Kobak et al, 2015 [38]; Macias et al, 2015 [40]; Meiland et al, 2012 [41]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Palmier-Claus et al, 2013 [46]; Prada et al, 2017 [47]; Rizvi et al, 2016 [48]; Rohatagi et al, 2016 [49]; Sands et al, 2016 [54]; Sze et al, 2015 [51]; Villalobos et al, 2017 [55]; Wood et al, 2017 [56]
Interview	Auger et al, 2014 [24]; Boman and Bartfai, 2015 [29]; Carey et al, 2016 [57]; Connelly et al, 2016 [58]; Corden et al, 2016 [31]; de Korte et al, 2018 [59]; Dulin et al, 2014 [33]; Fuller-Tyszkiewicz et al, 2018 [34]; Huguet et al, 2015 [36]; Kizakevich et al, 2018 [61]; Meiland et al, 2012 [41]; Mistler et al, 2017 [42]; Morland et al, 2016 [43]; Nicholson et al, 2018 [44]; Nitsch et al, 2016 [45]; Ospina-Pinillos et al, 2018 [53]; Palmier-Claus et al, 2013 [46]; Price et al, 2016 [62]; Rohatagi et al, 2016 [49]; Ruggiero et al, 2015 [50]; Sands et al, 2016 [54]; van Osch et al, 2015 [64]; Villalobos et al, 2017 [55]
Observation	Auger et al, 2014 [24]; Boyd et al, 2017 [30]; Henry et al, 2017 [35]; Meiland et al, 2012 [41]; Ospina-Pinillos et al, 2018 [53]; Price et al, 2016 [62]; van Osch et al, 2015 [64]
Think-aloud	Ben-Zeev et al, 2014 [28]; Latif et al, 2015 [39]; Nitsch et al, 2016 [45]; Ospina-Pinillos et al, 2018 [53]; van Osch et al, 2015 [64]; Whiteman et al, 2017 [52]
App-use generated data	Dulin et al, 2014 [33]; Garcia et al, 2017 [60]; Macias et al, 2015 [40]; Nicholson et al, 2018 [44]
Task-based evaluation	Ben-Zeev et al, 2013 [27]; Henry et al, 2017 [35]; Ospina-Pinillos et al, 2018 [53]; van Osch et al, 2015 [64]
Focus group	Connelly et al, 2016 [58]; Garcia et al, 2017 [60]; Ruggiero et al, 2015 [50]
Sensor data	Garcia et al, 2017 [60]
Cooperative design	Whiteman et al, 2017 [52]
Verbal probing	Whiteman et al, 2017 [52]
User feedback	Sands et al, 2016 [54]
Video recording	Price et al, 2016 [62]
Diary	Meiland et al, 2012 [41]
Wireframing/sketching	Ospina-Pinillos et al, 2018 [53]
Personas	VanHeerwaarden et al, 2018 [63]
Journey mapping	VanHeerwaarden et al, 2018 [63]

Table 8 lists the type of evaluations undertaken in our review, either formative or summative, according to the evaluator having a medical or computer science background obtained from author affiliations and biographies in the articles. A total of 3 papers reported both summative and formative usability evaluations as they reported on several phases of development. The most common occurrence was summative evaluations carried out by authors with a medical background. When computer scientists are involved in the usability evaluation, it is in collaboration with scientists with a medical background. There were no papers reporting a formative evaluation purely with authors who were computer scientists. We would expect the frequency of formative evaluations with computer scientists to be higher, as the goal of HCI research is improving technology, building on deep understanding of user perceptions and use patterns of technology. These observations could be explained by the table presenting the maturity of the technologies being evaluated and the evaluations being mostly concerned with matured and released versions of technologies.

Table 8. Author credentials category according to the evaluation type (N=42).

Type of evaluation	Medical credentials, studies (n)	Computer science credentials, studies (n)	Both, studies (n)
Formative evaluation	5	—[a]	9
Summative evaluation	17	3	11

[a]No study fulfills this criterion.

Discussion

Approach to the Users

According to the data yielded by the literature search, most of the studies conducted a usability evaluation of a mobile mental health program with diagnosed patients. It is possible to evaluate the usability of mHealth technologies with healthy users, but many of the studies in our review were simultaneous trials, with the goal of measuring health outcomes in addition to the effects of technology, which can explain the high number of studies that evaluated usability with patients. Although the evaluated programs were within the scope of the mental health domain, some of the studies recruited healthy users to measure usability and understand how to meet user expectations and needs. Studies gathering data from healthy users mainly followed a user-centered design approach, focusing on the development and evaluation process of a mobile app. For example, Connelly et al [58] developed a mobile app for low literacy to record unhealthy eating and weight control behaviors of Mexican American women. The authors completed the development process in 4 phases and conducted a final usability assessment. Similarly, Ospina-Pinillos et al [53] used participatory research methodologies to develop a mental health e-clinic for healthy young people across Australia. The authors included young people in all stages of the development process. It is interesting that the majority of the studies involved patients as it is much more difficult to recruit patients than healthy users, and for some diagnoses, such as dementia or schizophrenia, there may also be particular challenges in working with these patients to learn about the usability issues. In this respect, participation in the co-design and evaluation of technology to treat an illness also concerns the aspect of patient representation (eg, [67]); that is, those affected by a mental health problem should be able to influence the design of technology that is being made to treat and manage the problems.

Although most studies evaluated the usability of mobile mental health programs with a single user group, one-third enriched the usability data with different groups of potential users, such as patients, health care providers, healthy users, affected parents and children, and medical experts. For example, Boyd et al [30] involved both healthy users and patients, Ruggiero et al [50] included both health care providers and children, and Fuller-Tyszkiewicz et al [34] conducted a usability evaluation with health care providers, patients, and researchers in the health domain. The health care providers from whom the reviewed studies gathered data were clinicians, caregivers, nurses, therapists, care managers, or health professionals. When the goal of the mHealth technology is to change a medical practice, rather than improve health directly, it becomes important to involve other groups in the usability evaluations; for example, Boman and Bartfai [29] evaluated a physical robot as assistive technology for enhancing communication between patients and health care professionals. Each different user group or stakeholders may have different views on how useful it is and different types of use, which affects perceptions of usability. Fuller-Tyszkiewicz et al [34] tested the usability of a mobile app for depression with patients, mental health professionals, and researchers in the health domain. The WHO [5] identified health workers' perceived barriers to using mHealth apps, and one of the main barriers was usability problems in the apps and problems with integrating the new tools with systems already in use. There are several barriers to changing existing work practices for health workers, and for mobile devices, health workers are concerned about the character limits on SMS messages, and limited/cumbersome note-taking capabilities [5]. However, health workers are interested in being involved in the design and evaluation of new technology [5].

Although the studies in this review aimed to test the usability of a mobile mental health program, only half emphasized usability evaluation as the main purpose. A total of 10 studies highlighted feasibility ("an assessment of the practicality of a proposed plan or method," [68]) and acceptability ("the extent to which the assessment is experienced as probing yet unobtrusive," [66]) in addition to the usability evaluation. A few of the studies addressed the components of usability, such as effectiveness, user satisfaction, and efficiency. A considerable number of studies described the process of software development life cycle including design, development or adaptation, whereas the majority of the studies carried out a summative evaluation. Only 11 studies engaged in a formative evaluation to gather feedback from users and improve the design as part of an iterative design process. This is in line with the finding of Nørgaard and Hornbæk [12], who reported that the data from the evaluation of prototypes was rarely used in interaction design for reasons such as the lack of action ability of the evaluation results and time pressure in the development process. Kjeldskov and Stage [16] also pointed out that it was easier to carry out formative evaluations early in the development process, whereas there were stronger obstacles to changing the designs later. Only 12 of the reviewed studies were in the sketch or prototype stages, whereas 39 were matured or released versions of technology. When evaluating the usability of a finished technology, the goal becomes to demonstrate the effectiveness and validate the design rather than improving it.

Objectives and Outcomes

We found that many of the studies were heavily influenced by practices, ideas, and notions from randomized controlled trials, which is the standard practice for evidence building in medicine. Many of the studies set out to investigate the feasibility and acceptability of an mHealth app (eg, [28,32,42]), and usability

measures were used as a step toward fulfilling this goal. Feasibility and acceptability are often the focus of the pilot phase in a randomized controlled trial for a new medical procedure or medication. The word *usability* was not included in the title of most of the studies, although usability evaluation was either the main goal or one of the goals. Some of the reviewed studies (eg, [28,30,32,51,54,57]) also examined feasibility and/or acceptability but used the terms usability and acceptability together. For some of the studies (eg, [32,38,61]), the goal was to measure effectiveness of a mobile app for a specific mental health problem. They mainly administered a usability questionnaire as a summative evaluation at the end of their field study or trial. Furthermore, there were studies measuring only simplicity of use (eg, [30]), ease of use, and usefulness (eg, [59]). There was a duality in the goals for and underlying assumptions of developing a digital tool to be usable, that is, an mHealth app, and attempting to improve a person's health. In usability evaluations, the goal was to learn whether a tool is meaningful and how it could be improved, whereas in studies of health and medicine, the goal was to create a positive health effect for a person. When these goals are combined, the objective of the usability studies becomes explaining how the health effects are assisted or mediated by the mHealth app, and the usability evaluation evidence has a summative role.

Concerning the outcomes that the included studies presented, all but one referred to user reception as the main contribution of the study. This finding indicates that almost all the studies received positive feedback from their participants who found the evaluated tools useful. Positive medical outcome, tool improvement, app/tool, and design recommendations were other commonly reported results in the studies. These studies had conclusions about the positive medical effects of the evaluated tools on the participants and how the mobile mental health program was improved and accomplished based on user feedback and recommendations to researchers or practitioners on using similar technologies. We regard the reporting of medical outcomes in usability studies as a part of building evidence of the effectiveness of the mHealth technology. Some of the reviewed studies (eg, [58,64,65]) contained the development process of a mobile mental health app in detail, and the authors of these studies elaborated how a mobile app was improved following an iterative and incremental process based on user feedback. Whiteman et al [52] suggested that early involvement of users in the development process resulted in building a usable system. Similarly, Juengst et al [37] listed the lessons learned, such as the importance of a simple interface of a mobile app and effective communication between patients and health care providers. These results can be an important step on the journey from an idea for a mobile mental health intervention technology to its implementation and use in health care, for example, in warranting further research.

Reporting these kinds of outcomes can also be understood as an attempt to demonstrate that the mHealth technology works or does what it is intended to do, which is a common venture in HCI research. Klasnja et al [69] argued that electronic health technology evaluations should refrain from documenting behavioral changes, as behavioral change processes (1) are inherently complex, that is, subject to interconnected social, material-logistical, motivational, and circumstantial factors, and (2) need to have a very long–time frame to be of value. The problem of attrition or lack of sustained use has also been described as specific to mHealth apps (eg, [9]). Alternatively, the evaluation could focus directly on the underlying behavioral change strategies of the mHealth app to warrant or unwarrant further investigation of the medical efficacy of the app [69], for example, to determine whether a particular implementation of the strategy of self-monitoring the number of steps walked in a day actually increased the number of steps. Early in the design process of a technology for behavioral change, a deep understanding of the "how and why of the technology use by its target users should be a central goal for evaluations" according to Klasnja et al [69].

Characteristics of the Mobile Health Interventions

Most of the reviewed studies evaluated matured versions of mobile mental health apps; therefore, the app had been previously tested by either users or patients and updated based on their feedback. This was followed by the released version, which refers to an app which is downloadable from a platform, such as Apple Store or Google Play, prototype version, which means the app has a high-fidelity version for users to test its usability and functionality, and prototype-to-matured version. Only 15 of the included studies were described as user-oriented design methods of either user-centered design or participatory design; however, most studies described their methods as trials. Studies using a user-oriented design method most often carried out a formative evaluation, whereas most of the studies describing their methods as a trial engaged in a summative evaluation.

The most commonly referred to reason for developing mHealth apps for mental health was the availability of mobile devices to users, their popularity, and how people in general became accustomed to using them for various purposes, for example, by pointing out how mobile technologies were in a "process of technological acceleration" [25], and a "mobile device explosion" [60] and that "smartphone users are with their phones for all but 2 hours every day" [57]. This way of supporting the development of apps for mobile use is arguably generic and transferable to other areas of mobile technology use, such as games and social media. Simultaneously, through the proliferation of digital technology into the private spheres of the users [14,15], mobiles have partly facilitated a shift in the design and use of technology, paving the way for fields of research, such as mHealth. A number of the studies approached mobility by regarding and making use of the affordances and characteristics specific to mobile devices, such as opportunities for sensing data about the user and their contexts [24] and facilitating communication [29]. A total of 5 papers mentioned the potential for mobiles to support privacy, which is in agreement with the WHO [5] report that found privacy in stigmatized health conditions as one of the feasibility enhancers for patients.

Research Methods and Techniques of the Studies

User evaluation is an essential source of information to improve the usability of systems [70], aims to understand both positive and negative sides of an app, and provides valuable information

in this regard [71]. To gather user feedback and evaluate the usability of mHealth apps, the most common data collection technique utilized in the studies was questionnaire, followed by field study, interview, observation, think-aloud, and app-use generated data. This result corroborates the ideas of Holzinger [72], who pointed out that among several usability assessment methods, questionnaire, think aloud, and observation were the most commonly used methods.

Accordingly, the SUS was the most common standard questionnaire used by the studies in the review. This scale, developed by Brooke [73], aims to measure perceived usability and is one of the well-established and popular scales in the HCI field. Some of the studies constructed their own questionnaires to evaluate the usability of mobile mental health programs based on available standard questionnaire(s). The SUS, the Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire [74], and the Poststudy System Usability Questionnaire [75] were the most commonly used scales that studies used to create their own questionnaires. Almost half of the studies which used questionnaire as the main data collection technique either created a questionnaire based on one of those available or constructed their own questionnaires as available questionnaires did not entirely meet their needs in evaluating the usability of mobile mental health programs. Considering that the questionnaires were targeted toward desktop apps, our results unveil a need for a questionnaire focusing on testing the usability of mobile mental health apps. There can be good reason to tailor and adapt a standard questionnaire to a particular usability evaluation, but the authors are then expected to prove the reliability and validity of their questionnaire. Among the reviewed studies which either created new items or adapted a standard questionnaire, only Prada et al [47] provided a reliability score (Cronbach α=.88) of the questionnaire they developed.

Implications and Recommendations for Future Mobile Health Research and Usability Evaluation

Involve Patients and Health Care Professionals in Mobile Health Development

As people access mHealth apps to improve their health, publishers of mHealth apps have a responsibility to ensure the medical quality of their app. Currently, app providers have no formal responsibility to ensure and communicate medical evidence of their effectiveness. One aspect of ensuring this quality is building and evaluating apps in collaboration with health care professionals. Our review found that only in 11 of 42 studies were health care professionals involved in usability evaluation. Currently, it is the technology companies rather than hospitals, clinics, or doctors that are the most frequent publishers of health care apps [76], and there is a lack of involvement from health professionals in these apps [77,78].

Equally, there is a need to involve patients in the design and evaluation of mHealth apps, for example, to ensure the relevance of the apps and to obtain the experiences, beliefs, and preferences of the intended users. Most of the studies in our review of the literature involved patients; however, it remains an HCI challenge to develop ways in which to foster relevant contributions from often vulnerable patient user groups to complex design processes [79].

Standardize a Questionnaire for Mobile Health Apps

As in our review, Perez-Jover et al [80] found that usability evaluation practices in mHealth varied substantially. Accordingly, McFay et al [81] found a lack of best practices or standards for evaluating mHealth apps and behavioral change technologies. In a review, they found that self-developed, nonvalidated evaluation checklists were the most common evaluation method. The lack of validation casts doubt on the reliability of the results. In this review, we found that questionnaire was the most common data collection technique of the included studies; however, researchers either used standard questionnaires, such as SUS or USE, which were not specifically designed for the mental health domain, or adapted a standard questionnaire or developed a new one. Owing to the great variety of the questionnaires, there is a need to establish a common standardized usability questionnaire targeted specifically at mHealth mental health apps.

Foster Increased Collaboration Between Health Care and Computer Science Professionals in Mobile Health Development

Our review found that there was limited collaboration between computer science professionals and health care professionals in mHealth development. Many of the studies were carried out solely by health care professionals, and usability was evaluated in a summative manner. There is reason to believe that even closer collaboration between health care and computer science experts in the usability evaluation of mHealth apps will increase the quality of the evaluation interpretations, especially for formative evaluations.

Limitations

There are some limitations to this study. One of these is that the study was restricted to mobile technologies, whereas several mHealth intervention technologies are available on other platforms, such as the Web for PC, and how usability is evaluated for these technologies is also important. A second limitation to this review is that we did not download and test any of the mHealth apps referred to in the reviews. Reading about an app gives a different impression than interacting with the app itself and has consequences for how we perceive the following usability evaluation, potentially limiting our understanding of this work. A third limitation pertains to the division between academically driven mHealth apps and the much larger portion developed by the technology industry as reviewing the literature for usability evaluation practices through academic databases resulted in only finding academic studies, which, in turn, influenced the usability evaluation practices we observed. Therefore, we have less knowledge of the usability evaluation practices in industry.

Conclusions

Based on the call for evidence of their effectiveness in the plethora of mHealth intervention technologies, this study provides a detailed account of how evidence is being gathered in the form of usability evaluations from the perspective of

computer science and HCI, including how users feature in the evaluation, which study objectives and outcomes are stated, which research methods and techniques are used, and what the notion of mobility features is in mHealth apps. The most common reasons for developing mobile mental health apps provided in the studies were the availability of mobile devices to users, their popularity, and device affordances. Most studies described their methods as trials gathered data from a small sample size and carried out a summative evaluation using a single questionnaire, indicating that usability evaluation was not the main focus. The extent to which a mobile mental health intervention is able to meet expectations and needs was linked to the effectiveness, efficiency, and satisfaction of such programs and thereby its usability [82]. Evidence from this literature review also indicated that almost all studies received positive feedback from their participants who found the evaluated tools useful. However, further research is required to investigate the effects of usability levels of mobile mental health apps on outcomes of an intervention. As many of the studies described using an adapted version of a standard usability questionnaire, there is a need to develop a standardized mHealth usability questionnaire, which is a goal of future research.

Acknowledgments

This publication is part of the INTROducing Mental health through Adaptive Technology project, funded by the Norwegian Research Council (259293/o70).

Conflicts of Interest

None declared.

References

1. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
2. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
3. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015 May 19;6(1):27882. [doi: 10.3402/ejpt.v6.27882]
4. IMS Institute for Healthcare Informatics. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2019-06-05]
5. WHO Guideline: recommendations on digital interventions for health system strengthening. Geneva: World Health Organization; 2019.
6. Wykes T, Schueller S. Why reviewing apps is not enough: transparency for trust (t4t) principles of responsible health app marketplaces. J Med Internet Res 2019 May 02;21(5):e12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]
7. Lindgaard G. Early traces of usability as a science and as a profession. Interact Comput 2009 Dec;21(5-6):350-352. [doi: 10.1016/j.intcom.2009.03.006]
8. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
9. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676 [FREE Full text] [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
10. Nielsen J. Nielsen Norman Group. Usability 101: Introduction to Usability. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
11. Hornbæk K. Dogmas in the assessment of usability evaluation methods. Behav Inf Technol 2010 Jan;29(1):97-111. [doi: 10.1080/01449290801939400]
12. Nørgaard M, Hornbæk K. Exploring the value of usability feedback formats. Int J Hum Comput Interact 2009 Jan 14;25(1):49-74. [doi: 10.1080/10447310802546708]
13. Law EL, Abrahão S. Interplay between User Experience (UX) evaluation and system development. Int J Hum Comput Interact 2014 Jun;72(6):523-525. [doi: 10.1016/j.ijhcs.2014.03.003]
14. Bødker S. When second wave HCI meets third wave challenges. Oslo, Norway: Changing Roles; 2006 Presented at: Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles; October 14-18; Oslo, Norway p. 1-8. [doi: 10.1145/1182475.1182476]
15. Bødker S. Third-wave HCI, 10 years later---participation and sharing. In: Interactions. New York: ACM; Aug 28, 2015:24-31.
16. Kjeldskov J, Stage J. New techniques for usability evaluation of mobile systems. Int J Hum Comput Stud 2004 May;60(5-6):599-620. [doi: 10.1016/j.ijhcs.2003.11.001]

17. Brown B, Reeves S, Sherwood S. Into the wild: challenges and opportunities for field trial methods. 2011 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 07-12; Vancouver, BC, Canada p. 1657-1666. [doi: 10.1145/1978942.1979185]
18. Zhang D, Adipat B. Challenges, methodologies, and issues in the usability testing of mobile applications. Int J Hum Comput Interact 2005 Jul;18(3):293-308. [doi: 10.1207/s15327590ijhc1803_3]
19. Mantyjarvi J, Paternò F, Salvador Z, Santoro C. Scan and tilt: towards natural interaction for mobile museum guides. 2006 Presented at: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services; September 12-15; Helsinki, Finland p. 191-194. [doi: 10.1145/1152215.1152256]
20. Francone J, Nigay L. Using the user's point of view for interaction on mobile devices. 2011 Presented at: Proceedings of the 23rd Conference on l'Interaction Homme-Machine; October 24-27; Sophia Antipolis, France. [doi: 10.1145/2044354.2044360]
21. Butler A, Izadi S, Hodges S. SideSight: multi-"touch" interaction around small devices. In: Proceedings of the 21st annual ACM symposium on User interface software and technology. 2008 Presented at: 21st annual ACM symposium on User interface software and technology; October 19-22; Monterey, CA, USA p. 201-204. [doi: 10.1145/1449715.1449746]
22. Marsden G, Jones M. Mobile Interaction Design. USA: Wiley, John & Sons, Incorporated; 2006.
23. 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]
24. Auger C, Leduc E, Labbé D, Guay C, Fillion B, Bottari C, et al. Mobile applications for participation at the shopping mall: content analysis and usability for persons with physical disabilities and communication or cognitive limitations. Int J Environ Res Public Health 2014 Dec;11(12):12777-12794 [FREE Full text] [doi: 10.3390/ijerph111212777] [Medline: 25513999]
25. Barrio P, Ortega L, López H, Gual A. Self-management and shared decision-making in alcohol dependence via a mobile app: a pilot study. Int J Behav Med 2017 Oct;24(5):722-727. [doi: 10.1007/s12529-017-9643-6] [Medline: 28236288]
26. Bauer AM, Hodsdon S, Bechtel JM, Fortney JC. Applying the principles for digital development: case study of a smartphone app to support collaborative care for rural patients with posttraumatic stress disorder or bipolar disorder. J Med Internet Res 2018 Jun 06;20(6):e10048 [FREE Full text] [doi: 10.2196/10048] [Medline: 29875085]
27. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
28. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
29. Boman I, Bartfai A. The first step in using a robot in brain injury rehabilitation: patients' and health-care professionals' perspective. Disabil Rehabil Assist Technol 2015;10(5):365-370. [doi: 10.3109/17483107.2014.913712] [Medline: 24766268]
30. Boyd A, Synnott J, Nugent C, Elliott D, Kelly J. Community-based trials of mobile solutions for the detection and management of cognitive decline. Healthc Technol Lett 2017 Jun;4(3):93-96 [FREE Full text] [doi: 10.1049/htl.2016.0102] [Medline: 28706726]
31. Corden ME, Koucky EM, Brenner C, Palac HL, Soren A, Begale M, et al. MedLink: a mobile intervention to improve medication adherence and processes of care for treatment of depression in general medicine. Digit Health 2016;2:2055207616663069 [FREE Full text] [doi: 10.1177/2055207616663069] [Medline: 29942564]
32. Deady M, Johnston D, Milne D, Glozier N, Peters D, Calvo R, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 04;6(12):e11661 [FREE Full text] [doi: 10.2196/11661] [Medline: 30514694]
33. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
34. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
35. Henry JA, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone "App" for tinnitus management. Int J Audiol 2017 Oct;56(10):784-792. [doi: 10.1080/14992027.2017.1338762] [Medline: 28669224]
36. Huguet A, McGrath PJ, Wheaton M, Mackinnon SP, Rozario S, Tougas ME, et al. Testing the feasibility and psychometric properties of a mobile diary (myWHI) in adolescents and young adults with headaches. JMIR Mhealth Uhealth 2015 May 08;3(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3879] [Medline: 25956377]
37. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]

38. Kobak KA, Mundt JC, Kennard B. Integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Ann Gen Psychiatry 2015;14:37 [FREE Full text] [doi: 10.1186/s12991-015-0077-8] [Medline: 26535048]
39. Latif S, Tariq R, Tariq S, Latif R. Designing an assistive learning aid for writing acquisition: a challenge for children with dyslexia. Stud Health Technol Inform 2015;217:180-188. [Medline: 26294471]
40. Macias C, Panch T, Hicks YM, Scolnick JS, Weene DL, Öngür D, et al. Using smartphone apps to promote psychiatric and physical well-being. Psychiatr Q 2015 Dec;86(4):505-519. [doi: 10.1007/s11126-015-9337-7] [Medline: 25636496]
41. Meiland FJ, Bouman AI, Sävenstedt S, Bentvelzen S, Davies RJ, Mulvenna MD, et al. Usability of a new electronic assistive device for community-dwelling persons with mild dementia. Aging Ment Health 2012;16(5):584-591. [doi: 10.1080/13607863.2011.651433] [Medline: 22360649]
42. Mistler LA, Ben-Zeev D, Carpenter-Song E, Brunette MF, Friedman MJ. Mobile mindfulness intervention on an acute psychiatric unit: feasibility and acceptability study. JMIR Ment Health 2017 Aug 21;4(3):e34 [FREE Full text] [doi: 10.2196/mental.7717] [Medline: 28827214]
43. Morland LA, Niehaus J, Taft C, Marx BP, Menez U, Mackintosh M. Using a mobile application in the management of anger problems among veterans: a pilot study. Mil Med 2016 Sep;181(9):990-995. [doi: 10.7205/MILMED-D-15-00293] [Medline: 27612342]
44. Nicholson J, Wright SM, Carlisle AM, Sweeney MA, McHugo GJ. The WorkingWell mobile phone app for individuals with serious mental illnesses: proof-of-concept, mixed-methods feasibility study. JMIR Ment Health 2018 Oct 25;5(4):e11383 [FREE Full text] [doi: 10.2196/11383] [Medline: 30361199]
45. Nitsch M, Dimopoulos CN, Flaschberger E, Saffran K, Kruger JF, Garlock L, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972] [Medline: 26753539]
46. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
47. Prada P, Zamberg I, Bouillault G, Jimenez N, Zimmermann J, Hasler R, et al. EMOTEO: a smartphone application for monitoring and reducing aversive tension in borderline personality disorder patients, a pilot study. Perspect Psychiatr Care 2017 Oct;53(4):289-298. [doi: 10.1111/ppc.12178] [Medline: 27439663]
48. Rizvi SL, Hughes CD, Thomas MC. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
49. Rohatagi S, Profit D, Hatch A, Zhao C, Docherty JP, Peters-Strickland TS. Optimization of a digital medicine system in psychiatry. J Clin Psychiatry 2016 Sep;77(9):e1101-e1107 [FREE Full text] [doi: 10.4088/JCP.16m10693] [Medline: 27487251]
50. Ruggiero KJ, Bunnell BE, Andrews Iii AR, Davidson TM, Hanson RF, Danielson CK, et al. Development and pilot evaluation of a tablet-based application to improve quality of care in child mental health treatment. JMIR Res Protoc 2015 Dec 30;4(4):e143 [FREE Full text] [doi: 10.2196/resprot.4416] [Medline: 26717906]
51. Sze YY, Daniel TO, Kilanowski CK, Collins RL, Epstein LH. Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: a feasibility study. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e97 [FREE Full text] [doi: 10.2196/mhealth.4603] [Medline: 26678959]
52. Fortuna KL, Lohman MC, Gill LE, Bruce ML, Bartels SJ. Adapting a psychosocial intervention for smartphone delivery to middle-aged and older adults with serious mental illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828 [FREE Full text] [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]
53. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eclinic to improve access to and quality of mental health care for young people: Using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
54. Sands N, Elsom S, Keppich-Arnold S, Henderson K, King P, Bourke-Finn K, et al. Investigating the validity and usability of an interactive computer programme for assessing competence in telephone-based mental health triage. Int J Ment Health Nurs 2016 Feb;25(1):80-86. [doi: 10.1111/inm.12165] [Medline: 26365233]
55. Villalobos O, Lynch S, DeBlieck C, Summers L. Utilization of a mobile app to assess psychiatric patients with limited English proficiency. Hisp J Behav Sci 2017 May 09;39(3):369-380. [doi: 10.1177/0739986317707490]
56. Wood AE, Prins A, Bush NE, Hsia JF, Bourn LE, Earley MD, et al. Reduction of burnout in mental health care providers using the provider resilience mobile application. Community Ment Health J 2017 May;53(4):452-459. [doi: 10.1007/s10597-016-0076-5] [Medline: 28070775]
57. Carey TA, Haviland J, Tai SJ, Vanags T, Mansell W. MindSurf: a pilot study to assess the usability and acceptability of a smartphone app designed to promote contentment, wellbeing, and goal achievement. BMC Psychiatry 2016 Dec 12;16(1):442 [FREE Full text] [doi: 10.1186/s12888-016-1168-z] [Medline: 27955643]
58. Connelly K, Stein KF, Chaudry B, Trabold N. Development of an ecological momentary assessment mobile app for a low-literacy, mexican american population to collect disordered eating behaviors. JMIR Public Health Surveill 2016 Jul 14;2(2):e31 [FREE Full text] [doi: 10.2196/publichealth.5511] [Medline: 27418020]

59. de Korte EM, Wiezer N, Janssen J, Vink P, Kraaij W. Evaluating an mHealth app for health and well-being at work: mixed-method qualitative study. JMIR Mhealth Uhealth 2018 Mar 28;6(3):e72 [FREE Full text] [doi: 10.2196/mhealth.6335] [Medline: 29592846]
60. Plaza García I, Sánchez CM, Espílez Á, García-Magariño I, Guillén GA, García-Campayo J. Development and initial evaluation of a mobile application to help with mindfulness training and practice. Int J Med Inform 2017 Sep;105:59-67. [doi: 10.1016/j.ijmedinf.2017.05.018] [Medline: 28750912]
61. Kizakevich PN, Eckhoff R, Brown J, Tueller SJ, Weimer B, Bell S, et al. Phit for duty, a mobile application for stress reduction, sleep improvement, and alcohol moderation. Mil Med 2018 Mar 01;183(suppl_1):353-363. [doi: 10.1093/milmed/usx157] [Medline: 29635566]
62. Price M, Sawyer T, Harris M, Skalka C. Usability evaluation of a mobile monitoring system to assess symptoms after a traumatic injury: a mixed-methods study. JMIR Ment Health 2016 Jan 11;3(1):e3 [FREE Full text] [doi: 10.2196/mental.5023] [Medline: 26753673]
63. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (thought spot) with transition-aged youth in postsecondary settings: participatory design research. J Med Internet Res 2018 Mar 06;20(3):e79 [FREE Full text] [doi: 10.2196/jmir.8102] [Medline: 29510970]
64. van Osch M, Rövekamp AJ, Bergman-Agteres SN, Wijsman LW, Ooms SJ, Mooijaart SP, et al. User preferences and usability of iVitality: optimizing an innovative online research platform for home-based health monitoring. Patient Prefer Adherence 2015;9:857-867 [FREE Full text] [doi: 10.2147/PPA.S82510] [Medline: 26170635]
65. Jiam NT, Hoon AH, Hostetter CF, Khare MM. IIAM (important information about me): a patient portability profile app for adults, children and families with neurodevelopmental disabilities. Disabil Rehabil Assist Technol 2017 Aug;12(6):599-604. [doi: 10.1080/17483107.2016.1198435] [Medline: 27377240]
66. Slade M, Thornicroft G, Glover G. The feasibility of routine outcome measures in mental health. Soc Psychiatry Psychiatr Epidemiol 1999 May;34(5):243-249. [doi: 10.1007/s001270050139] [Medline: 10396165]
67. Bell J, Leong TH. Collaborative Futures: Co-Designing Research Methods for Younger People Living with Dementia. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 04-09; Glasgow, Scotland UK URL: https://doi.org/10.1145/3290605.3300582 [doi: 10.1145/3290605.3300582]
68. Lexico Dictionary. URL: https://www.lexico.com/en/definition/feasibility_study
69. Klasjna 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. 2007 Oct Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 07-12, 2011; Vancouver, BC, Canada p. 1693-1699. [doi: 10.1145/1978942.1979396]
70. Iacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. In: 2013 10th Working Conference on Mining Software Repositories (MSR). 1960 Presented at: 0th Working Conference on Mining Software Repositories (MSR); May 18-19, 2013; San Francisco, CA, USA p. 103-104. [doi: 10.1109/MSR.2013.6624001]
71. Hamborg K, Vehse B, Bludau HB. Questionnaire based usability evaluation of hospital information systems. Electron J Inform Sys Eval 2004;7(1):21-30 [FREE Full text]
72. Holzinger A. Usability engineering methods for software developers. In: Communications of the ACM - Interaction Design and Children. New York: ACM; Jan 01, 2005:71-74.
73. Brooke J. SUS: a quickdirty usability scale. In: Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996:189-194.
74. Lund A. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
75. Lewis JR. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016 Aug 06 Presented at: Proc Hum Factors Ergon Soc Annu Meet ;36?60; October 1, 1992; Atlanta, GA p. 1259-1260. [doi: 10.1177/154193129203601617]
76. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
77. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
78. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 09;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
79. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: mixed methods study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
80. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]

81. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
82. Goodwin J, Cummins J, Behan L, O'Brien SM. Development of a mental health smartphone app: perspectives of mental health service users. J Ment Health 2016 Oct;25(5):434-440. [doi: 10.3109/09638237.2015.1124392] [Medline: 26732242]

Abbreviations

HCI: human-computer interaction
mHealth: mobile health
SUS: System Usability Scale
USE: Usefulness, Satisfaction, and Ease of Use
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.07.19; peer-reviewed by JJ Miranda, F Mckay; comments to author 29.09.19; revised version received 22.10.19; accepted 23.10.19; published 06.01.20.

Please cite as:
Inal Y, Wake JD, Guribye F, Nordgreen T
Usability Evaluations of Mobile Mental Health Technologies: Systematic Review
J Med Internet Res 2020;22(1):e15337
URL: https://www.jmir.org/2020/1/e15337
doi:10.2196/15337
PMID:31904579

Original Paper

Effectiveness and Cost-Effectiveness of a Self-Guided Internet Intervention for Social Anxiety Symptoms in a General Population Sample: Randomized Controlled Trial

John Powell[1], PhD; Veronika Williams[2], PhD; Helen Atherton[3], PhD; Kylie Bennett[4], PhD; Yaling Yang[1], PhD; Mina Davoudianfar[1], BA; Annika Hellsing[5], BPsyc; Angela Martin[1], PhD; Jill Mollison[1], PhD; Milensu Shanyinde[1], MSc; Ly-Mee Yu[1], PhD; Kathleen M Griffiths[5], PhD

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]School of Nursing, Faculty of Education and Professional Studies, Nipissing University, North Bay, ON, Canada

[3]Unit of Academic Primary Care, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[4]eHub Health Pty Ltd, Goulburn NSW, Australia

[5]Research School of Psychology, Australian National University, Canberra, Australia

Corresponding Author:
John Powell, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter, Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865617768
Email: john.powell@phc.ox.ac.uk

Abstract

Background: Many people are accessing digital self-help for mental health problems, often with little evidence of effectiveness. Social anxiety is one of the most common sources of mental distress in the population, and many people with symptoms do not seek help for what represents a significant public health problem.

Objective: This study aimed to evaluate the effectiveness of a self-guided cognitive behavioral internet intervention for people with social anxiety symptoms in the general population.

Methods: We conducted a two-group randomized controlled trial in England between May 11, 2016, and June 27, 2018. Adults with social anxiety symptoms who were not receiving treatment for social anxiety were recruited using online advertisements. All participants had unrestricted access to usual care and were randomized in a 1:1 ratio to either a Web-based unguided self-help intervention based on cognitive behavioral principles or a waiting list control group. All outcomes were collected through self-report online questionnaires. The primary outcome was the change in 17-item Social Phobia Inventory (SPIN-17) score from baseline to 6 weeks using a linear mixed-effect model that used data from all time points (6 weeks, 3 months, 6 months, and 12 months).

Results: A total of 2122 participants were randomized, and 6 were excluded from analyses because they were ineligible. Of the 2116 eligible randomized participants (mean age 37 years; 80.24%, 1698/2116 women), 70.13% (1484/2116) had follow-up data available for analysis, and 56.95% (1205/2116) had data on the primary outcome, although attrition was higher in the intervention arm. At 6 weeks, the mean (95% CI) adjusted difference in change in SPIN-17 score in the intervention group compared with control was −1.94 (−3.13 to −0.75; *P*=.001), a standardized mean difference effect size of 0.2. The improvement was maintained at 12 months. Given the high dropout rate, sensitivity analyses explored missing data assumptions, with results that were consistent with those of the primary analysis. The economic evaluation demonstrated cost-effectiveness with a small health status benefit and a reduction in health service utilization.

Conclusions: For people with social anxiety symptoms who are not receiving other forms of help, this study suggests that the use of an online self-help tool based on cognitive behavioral principles can provide a small improvement in social anxiety symptoms compared with no intervention, although dropout rates were high.

Trial Registration: ClinicalTrials.gov NCT02451878; https://clinicaltrials.gov/ct2/show/NCT02451878

(J Med Internet Res 2020;22(1):e16804) doi:10.2196/16804

KEYWORDS

randomized controlled trial; internet; self-care; social anxiety

Introduction

Background

Many people are accessing digital tools for self-help for a range of mental health problems [1]. Social anxiety is one of the most common sources of mental distress in the population and represents a significant public health problem [2]. It is characterized by a cluster of cognitive, behavioral, and physiological symptoms including an intense and persistent fear of being negatively evaluated in social or performance situations, along with avoidance of such situations. The individual fears that they may act in such a way, or show anxiety symptoms, which would lead to embarrassment or humiliation. A diagnosis of social anxiety disorder may be made when symptoms are persistent and lead to disruption of daily routine, and work or social life, or if the symptoms themselves cause marked distress. There is a spectrum of symptomatology in the general population, and even subclinical symptoms that do not reach a clinical diagnostic threshold can cause substantial impairment [3,4].

Effective psychological and pharmacological treatments exist for social anxiety symptoms, but many people with symptoms do not seek or receive these treatments [5-7]. Self-guided digital tools have received much attention owing to their potential for high scalability and low marginal cost, in addition to the benefits of convenient access and anonymity they offer to people with social anxiety symptoms who may not seek help through more traditional routes because of embarrassment or fear of scrutiny [8]. A 2014 meta-analysis of randomized controlled trials of unguided internet-based self-help for social anxiety disorder identified 5 studies (270 participants in total) showing evidence for effectiveness for these interventions with a pooled standardized mean difference of 0.66 (95% CI 0.39-0.94) [5]. Subsequent trials using self-help interventions that use cognitive behavioral approaches have found similar effect sizes (between-group effect sizes ranging from 0.47 to 0.76) [9-14]. Previous studies were conducted on a relatively small scale (the largest number of participants in the intervention group in any previous individual study was 100 [14]) and have generally been confined to cases of social anxiety of clinical severity, usually based on a structured interview assessment. Very little work has attempted to examine the value of unguided self-help in a real-world context, where individuals self-select as requiring help with symptoms that may not reach a clinical threshold but may be causing them some level of distress and choose to access digital tools themselves, with no clinician contact at all. In this study, we examine the effectiveness and cost-effectiveness of the self-help E-couch social anxiety tool (described in detail below). This was chosen as it is a self-directed online intervention based on cognitive behavioral therapy principles including components of known effectiveness in face-to-face therapy. A previous laboratory-based comparative study of the E-couch social anxiety tool with 21 participants (mainly university students) in the E-couch arm showed pretest to posttest improvement in social anxiety measures [15].

Objectives

We, therefore, undertook the first large-scale pragmatic randomized trial of an online self-guided cognitive behavioral intervention for people with self-reported social anxiety symptoms in the general population. Our experimental hypothesis was that participants who received the intervention would have a greater improvement in symptoms of social anxiety compared with participants who did not.

Methods

Trial Design and Participants

A two-arm, parallel-group randomized controlled trial was conducted to compare the effectiveness and cost-effectiveness of a Web-based and mobile-optimized self-guided intervention with a waiting list control condition for treating social anxiety symptoms. The study received ethics approval from the University of Oxford Medical Sciences Inter-Divisional Research Ethics Committee (MS-IDREC-C1-2015-167) and the Australian National University Human Research Ethics Committee (Protocol 2015/229) and is registered on ClinicalTrials.gov (NCT02451878). All participants provided informed consent to take part in the study using a self-completion online form. Outcomes were assessed at baseline, 6 weeks, 3 months, 6 months, and 12 months. The main follow-up time points were chosen to measure immediate effect (6 weeks as the intervention was designed to be used over a 6-week period) and long-term outcomes (12 months), along with interim time points (3 months and 6 months) to strengthen our repeated measures analysis and to support participant engagement. All study administration was conducted using automated online systems. The trial protocol is in Multimedia Appendix 1.

Participants were recruited primarily through an online advertisement placed on the UK National Health Service (NHS) website. In addition, study advertisements seeking individuals with social anxiety symptoms were placed on university and charity websites and disseminated via email and social media. We aimed to capture people with a broad range of social anxiety symptoms in the general population, who were likely to be typical of those seeking help from self-directed digital tools. Interested potential participants completed an online screening questionnaire to assess eligibility. We excluded anyone currently receiving therapist-guided treatment for social anxiety disorder or who self-reported a diagnosis of schizophrenia or bipolar affective disorder. Initial inclusion criteria were having access to the internet-based intervention, aged 18 years or older, resident in England, having an email address and mobile telephone number (to receive study emails and text alerts), and an initial criterion of scoring in a subclinical range of 13 to 19 on the 17-item Social Phobia Inventory (SPIN-17). We had

initially chosen the 13 to 19 range with expert advice as this would, in theory, capture those scoring above the population mean score (11-12) while excluding those scoring above the commonly used threshold of 19, which indicates further assessment may be warranted (although this threshold does not represent a diagnosis). However, early in recruitment, it became apparent that most people in the general population volunteering for this study scored much higher than this, and the distribution of SPIN-17 scores meant that very few scored in the low range. There was clear evidence of a high level of unmet need among individuals living with social anxiety symptoms in the community and not seeking help elsewhere. With advice from our independent Trial Steering Committee, we, therefore, revised and reregistered the protocol (in line with good practice in clinical trials) to modify the inclusion criteria to include all individuals scoring 13 or more on SPIN-17, therefore capturing those in our hypothesized subclinical range of 13 to 19, as well as those with a higher score. We continued to exclude anyone receiving professional help, and therefore, the final sample represented adults in the general population who self-reported some level of social anxiety symptoms but who were not receiving treatment for social anxiety. Potentially eligible participants completed consent, and 24 hours later, they were sent an email link to self-complete their baseline measures using online questionnaires.

Randomization and Masking

Once baseline measures had been completed, participants were randomized (1:1 ratio) to either the intervention group (E-couch) or the waiting list control group using a computer-generated random number sequence run through an automatic online program using a block size of 2 without stratification. Due to the nature of the intervention, participants were not blind to allocation.

Interventions

Given that this was a general population sample, all participants continued to receive usual care. Participants in the intervention arm were given access to a password-protected website that contained the E-couch social anxiety module. The website was mobile-optimized and could, therefore, be used on a smartphone with the look and feel of a dedicated app, or on a computer browser.

Self-Guided Intervention

The E-couch social anxiety module is a self-directed interactive program based on cognitive behavioral therapy principles. The program is divided into 6 modules: a literacy section, which provides information about the symptoms of social anxiety, types of available help, and effective treatments, and 5 toolkits comprising exposure practice, cognitive restructuring (modifying your thinking), attention practice, social skills training, and relaxation. The content of the toolkits consists of evidence-based information, interactive exercises, and workbooks based on cognitive behavioral principles; participants could complete the modules in any order. Participants were advised to access and use the intervention over the initial period of 6 weeks (although they could work through the intervention at their own pace and were able to access it for the full 12-month duration of the study). "Ideal" usage of the intervention would entail engagement with 1 new module each week and ongoing updates to diaries and workbooks based on the user's real-life experiences. E-couch was developed by the ehub team at the Australian National University National Institute for Mental Health Research. The intervention was adapted for this study to create a "stand-alone" social anxiety intervention that was accessed via a password-protected portal and with the usual E-couch branding removed. The program was adapted for a UK audience by removing Australian-specific terminology and undertaking user testing on the new version. No changes were made to the intervention during the study period.

Waiting List Control

Participants in the control group were informed that they had been put on a waiting list to receive access to the intervention in 12 months. They were asked to complete baseline and follow-up measures at the same time as participants in the intervention group. They received no other intervention.

Automated text (SMS) message and email reminders were sent to participants in both groups to reduce attrition. Participants in the intervention condition received 1 text message within 24 hours of randomization to remind them to access the intervention and 3 email reminders during the 6-week intervention period to remind them to log in to access the program. In addition, all participants received email invitations to complete follow-up surveys at each outcome measure time point, with those who failed to complete receiving a reminder email followed by a reminder SMS text message.

Outcomes

The primary outcome was the change in SPIN-17 score from baseline to 6 weeks. The SPIN-17 is a 17-item self-rated scale covering the main social anxiety symptoms of fear, avoidance, and physiological disturbance. The responses to 17 statements (such as "I avoid talking to people I don't know") are rated on a 5-point scale from "not at all" (score=0) to "extremely" (score=4) to indicate the extent to which each statement reflects how the respondent was feeling in the past week, with higher scores reflecting greater social anxiety symptoms. The SPIN-17 has good test-retest reliability, internal consistency, and convergent and divergent validity [16]. Secondary outcomes were all also self-report measures with good reliability and validity: the 8-item Brief Fear of Negative Evaluation (BFNE-S scale), which is very commonly used in studies of social anxiety and measures one of the key psychological constructs of social anxiety (example item: "I am frequently afraid of other people noticing my shortcomings") [17]; the 20-item Centre for Epidemiologic Studies Depression scale (CES-D), which has been widely used in online studies of anxiety and depression interventions to measure depressive symptoms [18]; the 7-item Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS), a measure of mental well-being requiring participants to provide the extent of their agreement with statements about thoughts and feelings over the previous 2 weeks, which has been shown to be responsive to change (example item: "I've been feeling useful") [19,20]; and the widely used and validated 36-item Short Form Health Survey (version 1) to measure health status and quality of life expressed

in mental and physical component scores [21]. We also measured usage of the intervention in terms of number of E-couch modules completed, total time in minutes spent on modules, and total page views. Adverse events were not anticipated, but participants were asked to self-report any ill effects thought to be related to the intervention.

Sample Size

We aimed to recruit 2104 participants (ie, 1052 per group) to this trial, to detect a small between-group standardized mean difference of 0.2 at 5% two-sided significance level and 90% power, assuming a high level of potential attrition of up to 50% given the fully self-guided nature of the intervention and automated nature of the trial (all trial procedures were conducted online). Although previous studies have suggested a larger treatment effect for internet-delivered interventions, we believed this treatment effect is too optimistic for a pragmatic trial of a self-guided treatment in a general population sample. The target effect size, although small at an individual level, can potentially translate into an important population-level change [22].

Statistical Analyses

The statistical analysis was finalized before unblinding of the data. Primary analysis was modified intention to treat according to allocated group irrespective of adherence and with at least one outcome questionnaire completed post randomization. A linear mixed-effect model was fitted to the primary outcome data, using data collected at 6 weeks, 3 months, 6 months, and 12 months. Participant was included as a random intercept. Randomized group, baseline SPIN-17 score, time, and time by randomized group interaction term were fitted as fixed effects. An unstructured variance covariance matrix was specified between repeated measures on the same individual. Assumptions of normality and constant variance for linear mixed-effects models were assessed by residual plots and other diagnostics plots.

Given the high level of attrition expected in online trials of self-guided digital interventions, we also prespecified a Complier Average Causal Effect (CACE) analysis for the primary outcome and 2 other main outcomes, to include only participants who completed at least one module of the intervention and at least one outcome assessment to investigate the effect of the intervention in participants who adhered to the intervention. An instrumental variable approach was adopted to provide the CACE estimate at 6 weeks. This method involved a 2 least squares (using the “ivregress 2sls” command in STATA SE Version 15.1, StataCorp, Texas) estimation from fitting a linear regression model of the primary outcome, adjusting for baseline SPIN-17 and compliance instrumented on randomized group [23].

Similar approaches were undertaken for other outcomes. A CACE analysis at 6 weeks was conducted on fear of negative evaluation (BFNE-S) and mental well-being (SWEMWBS) measures, similarly adjusting for baseline BFNE-S and SWEMWBS in the linear regression models. Safety analyses were not conducted as there were no adverse events reported during the study period.

In anticipation of high levels of dropout, we prespecified various sensitivity analyses to explore the impact of assumptions regarding missing data in the primary outcome analysis. These included analyses (1) of participants with complete data at all time periods, (2) adjusting for factors found to be predictive of missingness, (3) fitting a pattern mixture model to assess different degrees of missing not at random, as well as (4) an assessment of missing not at random assumption for the primary outcome by assuming plausible arm-specific differences of missing SPIN-17 score between responders (with SPIN-17 score at 6 weeks) and nonresponders (missing SPIN-17 score at 6 weeks) [24,25].

Predefined subgroup analyses were conducted on change at 6 weeks for SPIN-17, BFNE-S, CES-D, and SWEMWBS for baseline SPIN-17 (<19, ≥19) to ascertain if the benefit differed between groups scoring above or below the screening threshold and for certain demographic characteristics to determine if the effectiveness of the intervention differed by the individual characteristics we had measured, that is, age (≤35, >35 years), gender (male, female), educational level (degree, no degree), and ethnicity (any white, nonwhite). Subgroup analyses were conducted by inclusion of an interaction term of baseline subgroup by randomized group by time in the linear mixed model. Descriptive statistics were used to describe usage data, adherence, and self-reports of other help received during the study period. All statistical analyses were performed using STATA SE version 15.1 [26].

Economic Evaluation

A cost utility analysis from an NHS and social care perspective was conducted within this trial to assess the cost-effectiveness of the intervention. The total costs of developing, modifying, delivering, and maintaining the intervention were obtained, and the mean intervention cost was estimated for the participants recruited in the intervention group. Data on health care service utilization (for any reasons) were collected for all participants, including primary care consultations, hospital outpatient appointments, and hospital admissions. Unit costs for these health services were obtained from the Personal Social Services Research Unit (2016-2017) using national average costs [27]. Maximum follow-up was 1 year; therefore, no discounting was applied. The total and mean costs for the intervention and the waiting list control group were calculated. Effectiveness was measured in quality-adjusted life years (QALYs) using the under-the-curve approach by combining the duration of follow-up with the health status utilities at the start and end points. Health status was measured using the self-reported SF-36 measure at baseline, 6 weeks, 3 months, 6 months, and 12 months. The analysis examined short-term (6 weeks) and long-term (12 months) impact. Health status utilities were converted from SF-36 to SF-6D indices using the established UK-based utility algorithm obtained through the University of Sheffield Licensing [28]. The primary outcome was the incremental cost per QALY gained between the intervention group at 6 weeks and 12 months.

Results

Participant Characteristics and Trial Flow

Recruitment took place between May 11, 2016, and May 9, 2017, when the target sample size was reached. Participants were followed up for 1 year. Final data were locked on June 27, 2018 (allowing time for delayed 12-month follow-up responses). Figure 1 shows the flow diagram of the participants throughout the study period. We screened 9447 participants of whom 5932 (62.79%) were ineligible, and a further 1393 (14.74%) did not complete the baseline measures. We randomized 1061 (1061/2122, 50.00%) participants to E-couch and 1061 (1061/2122, 50.00%) to the control group. A total of 6 participants who were randomized to the study were excluded from all analyses because their later responses indicated they did not meet the inclusion criteria in terms of age, leaving 2116 participants randomized and included in analyses. Table 1 shows the baseline characteristics, which were similar across both groups. Owing to a software error, many participants were not sent the email requesting completion of their interim (3 months or 6 months) outcome measures. This error did not affect emails sent at the main follow-up time points of baseline, 6 weeks, and 12 months, and data from all time points were included in the analysis. Attrition rates differed significantly between groups with an overall loss to follow-up at the main follow-up time point (6 weeks) of 42.9%, with a loss of 60.8% in the intervention arm and 25.3% in the control arm (see Figure 1 and Multimedia Appendix 2). By 12 months, the primary outcome was available for 349 of 1061 (32.89%) participants in the intervention group and 710 of 1061 participants (66.92%) in the control group.

Figure 1. Consolidated Standards of Reporting Trials flow diagram.

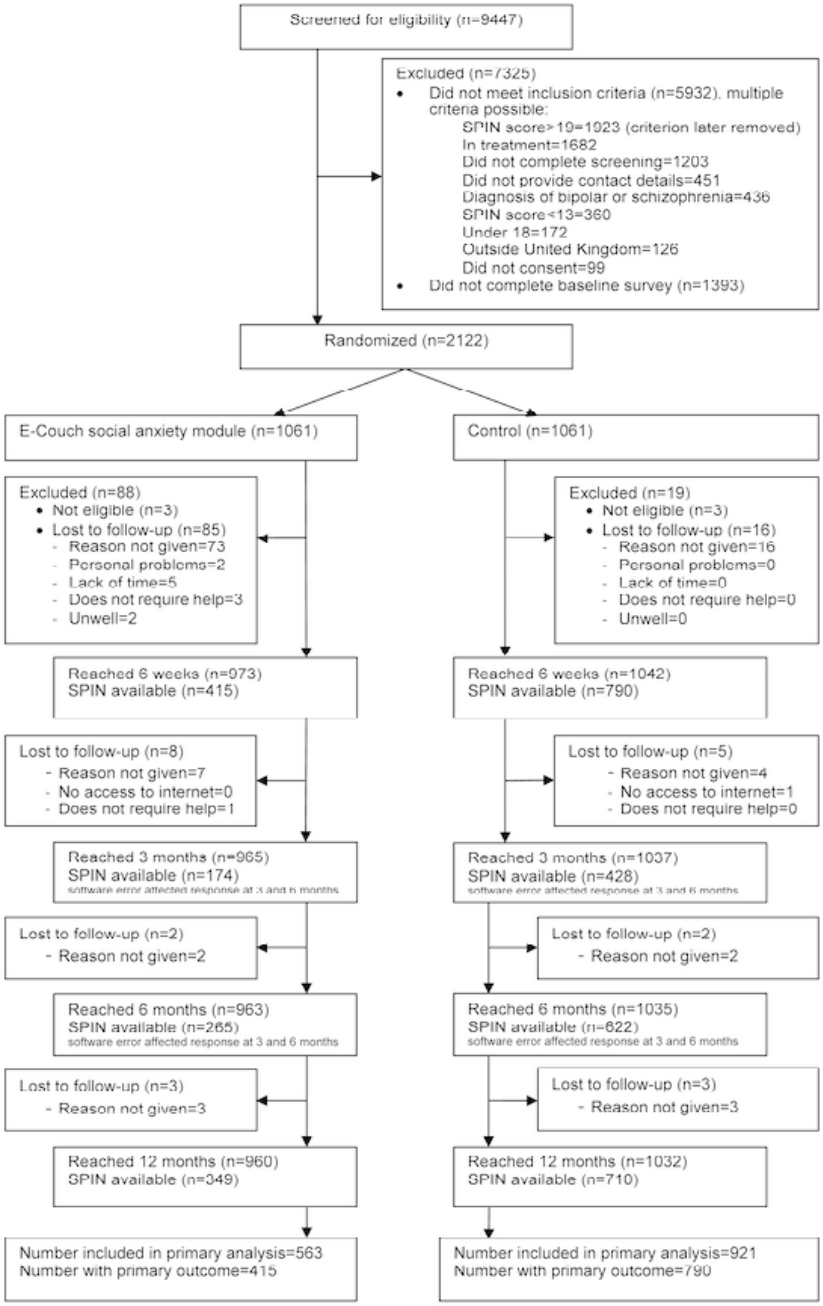


Table 1. Baseline characteristics of all participants randomized.

Baseline characteristics	E-couch (n=1058)	Control (n=1058)	Total randomized (N=2116)
Age (years), mean (SD); range	37.4 (13.9); 18-84	36.9 (13.6); 18-82	37.17 (13.8); 18-84
Gender, n (%)			
Female	859 (81.19)	839 (79.30)	1698 (80.25)
Male	187 (17.67)	213 (20.13)	400 (18.90)
Other	10 (0.95)	6 (0.57)	16 (0.76)
Missing	2 (0.19)	0 (0.00)	2 (0.09)
Marital status, n (%)			
Married or in a civil partnership	327 (30.91)	311 (29.40)	638 (30.15)
Not married	727 (68.71)	745 (70.42)	1472 (69.57)
Missing	4 (0.38)	2 (0.19)	6 (0.28)
Education, n (%)			
Degree	514 (48.58)	538 (50.85)	1052 (49.72)
No degree	544 (51.42)	520 (49.15)	1064 (50.28)
Employment status, n (%)			
Employed	610 (57.66)	623 (58.88)	1233 (58.27)
Unemployed	440 (41.59)	426 (40.26)	866 (40.93)
Missing	8 (0.76)	9 (0.85)	17 (0.80)
Income (£), n (%)			
≤25,000	736 (69.57)	739 (69.85)	1475 (69.71)
>25,000	322 (30.43)	319 (30.15)	641 (30.29)
Ethnicity, n (%)			
White	917 (86.67)	918 (86.77)	1835 (86.72)
Nonwhite	141 (13.33)	140 (13.23)	281 (13.28)
Social Phobia Inventory-17, mean (SD); range	39.6 (13.1); 13-68	39.8 (13.4); 13-68	39.7 (13.3); 13-68
Mental well-being score (Short Warwick-Edinburgh Mental Well-Being Scale), mean (SD); range	17.7 (3.1); 7.0-30.7	17.7 (3.3); 7.0-35.0	17.7 (3.2); 7.0-35.0
Brief Fear of Negative Evaluation score, mean (SD); range	22.5 (7.1); 0-32	22.4 (7.3); 0-32	22.4 (7.2); 0-32
Centre for Epidemiologic Studies Depression scale, mean (SD); range	30.5 (12.3); 2-60	30.7 (12.2); 0-58	30.6 (12.2); 0-60
Short Form-36 (physical component summary), mean (SD); range	50.2 (10.3); 12.2-69.3	49.8 (10.6); 49.8 (10.6)	50.0 (10.4); 12.2-69.3
Short Form-36 (mental component summary), mean (SD); range	49.9 (9.5); 29.3-77.1	50.1 (9.4); 30.8-83.9	50.0 (9.5); 29.3-83.9

Primary Outcome

Over the study period, there was a reduction of social anxiety symptoms in the E-couch group compared with that in the control group (see Multimedia Appendix 3). At 6 weeks, the E-couch group had a mean (SD) reduction of SPIN-17 score of −6.2 (10.8) and the control group −3.99 (9.3). The adjusted mean difference (95% CI; *P* value) in change in SPIN-17 score in E-couch compared with control was −1.94 (−3.13 to −0.75; *P*=.001; Table 2). This equates to a standardized mean difference effect size (between groups) of 0.2 (the pooled SD for SPIN-17 change was 9.81). At the 6-week follow-up, SPIN-17 outcome measures were available for 415 (415/1064, 39.00%) and 790 participants (790/1064, 74.25%) in the E-couch and control groups, respectively. In the CACE analysis, adjusted mean difference (95% CI; *P* value) in change in SPIN-17 score for intervention compared with control was −2.95 (−4.30 to −1.61; *P*<.001; Table 2). The results from the sensitivity analyses undertaken to explore missing data assumptions were also consistent with the primary outcome 6-week findings. These included analyses that only considered data from completers (defined as participants who returned all their outcome measures at the main time points of baseline, 6 weeks, and 12 months; see Multimedia Appendix 2) and the findings of the pattern mixture model even when assuming different missing data patterns in the intervention or control group. Finally, findings were also similar to the primary outcome analysis even under the assumption of missing not at random.

Table 2. Adjusted estimates from mixed-effect model for each outcome at 6 weeks and 12 months and estimates from the complier average causal effect analysis at 6 weeks.

Outcomes	Mixed-effect model analysis		Complier average causal effect analysis	
	Adjusted difference in mean change (95% CI)	*P* value	Adjusted difference in mean change (95% CI)	*P* value
Social Phobia Inventory-17[a]				
E-couch vs control (6 weeks)	−1.94 (−3.13 to −0.75)	.001	−2.95 (−4.30 to −1.61)	<.001
E-couch vs control (12 months)	−3.07 (−4.32 to −1.82)	<.001	N/A[b]	N/A
Brief Fear of Negative Evaluation score				
E-couch vs control (6 weeks)	−1.09 (−1.79 to −0.38)	.003	−1.60 (−2.38 to −0.82)	<.001
E-couch vs control (12 months)	−2.33 (−3.08 to −1.58)	<.001	N/A	N/A
Short Warwick-Edinburgh Mental Well-Being Scale				
E-couch vs control (6 weeks)	0.38 (−0.02 to 0.77)	.06	0.59 (0.17 to 1.02)	.006
E-couch vs control (12 months)	0.82 (0.39 to 1.24)	.001	N/A	N/A
Centre for Epidemiologic Studies Depression scale				
E-couch vs control (6 weeks)	−3.35 (−4.54 to −2.15)	<.001	N/A	N/A
E-couch vs control (12 months)	−1.79 (−3.06 to −0.52)	.006	N/A	N/A
Short Form-36 (physical component summary)				
E-couch vs control (6 weeks)	0.398 (−0.41 to 1.20)	.33	N/A	N/A
E-couch vs control (12 months)	0.003 (−0.85 to 0.85)	.99	N/A	N/A
Short Form-36 (mental component summary)				
E-couch vs control (6 weeks)	1.06 (0.12 to 1.98)	.03	N/A	N/A
E-couch vs control (12 months)	2.06 (1.07 to 3.06)	<.001	N/A	N/A

[a]Primary outcome.

[b]N/A: not applicable.

Secondary Outcomes

Table 2 shows the results for the secondary outcomes. At the 12-month follow-up, participants randomized to the E-couch group continued to show a greater reduction in severity of social anxiety symptoms than the control participants, with a mean (95% CI; *P* value) adjusted difference in change in SPIN-17 score of −3.07 (−4.32 to −1.82; *P*<.001; Table 2). As with the primary outcome, the results of the sensitivity analyses exploring missing data assumptions were consistent with the main analysis SPIN-17 findings at 12 months (Multimedia Appendix 2). The findings for the other outcome measures of fear of negative evaluation (BFNE-S), mental well-being (SWEMWBS), depression (CES-D), and the mental component scale of the SF-36 all showed statistically significant small improvements favoring E-couch compared with control (see Table 2). There was no evidence of difference between groups for the physical component scale of the SF-36. All distributions of residuals from the fitted models satisfied the normality assumption. No adverse events were reported during the study period.

Usage Data and Subgroup Analyses

At 6 weeks, the mean (SD) number of E-couch modules fully completed (out of 6) was 1.87 (1.43), total mean (SD) time in minutes spent on modules was 35.3 (48.1), and total mean (SD) page views was 37.6 (41.3). Greater adherence to the intervention was not associated with baseline SPIN-17 score, age, gender, or ethnicity (Multimedia Appendix 2). At 6 weeks, higher total page views or longer duration spent on modules was associated with larger improvement in social anxiety symptoms (Multimedia Appendix 2). At 6 weeks, E-couch had a significantly greater impact in improving social anxiety symptoms for participants with baseline SPIN-17 score greater than 19 (usually taken as cutoff to indicate clinical assessment warranted) compared with the few participants scoring in the lower range (SPIN-17 score 13-19; *P*=.01; Multimedia Appendix 2). In this subgroup analysis, the lower SPIN-17 scorers (the ones we had originally defined as a subclinical population) had no benefit from the intervention compared with the control group. The E-couch intervention also had a significantly greater beneficial impact on depressive symptoms at 6 weeks in participants with higher baseline SPIN-17 scores (*P*=.007), and again in this subgroup analysis, the few participants scoring in the lower SPIN-17 range had no benefit on depressive symptoms compared with the control group. There was no evidence of heterogeneity in the effects of intervention for the subgroup analyses involving BFNE-S and SWEMWBS.

Economic Evaluation

At both 6-week and 12-month follow-ups, the waiting list control group, in general, used more health care services than the E-couch group (see Tables 3 and 4). This resulted in a mean

health care cost saving of £26.48 at the 6-week follow-up and £65.04 at the 12-month follow-up. Adding the mean intervention cost of £48.40 to the intervention group, the E-couch group cost more than the control group at 6 weeks but is cost saving at 12 months. In the cost utility analysis, at both 6-week and 12-month follow-ups, there were very small improvements of general health status in both the E-couch group and the waiting list control group, with the E-couch group improvement slightly more than the control group: the SF-6D indices increased from 0.6 at baseline to 0.64 at 6 weeks and 0.66 at 12 months for the E-couch group and from 0.6 at baseline to 0.62 at 6 weeks and 0.64 at 12 months for the waiting list control group (see Table 5). At the 6-week follow-up, mean QALYs were 0.072 for the E-couch group and 0.070 for the control group, giving very small QALYs gains of 0.002 for the intervention over the control group. At the 12-month follow-up, mean QALYs were 0.635 for the E-couch group and 0.621 for the control group, with QALYs gain of 0.024 for the intervention over the control group. The incremental cost per QALY gained at 6 weeks was £10,960, which is highly likely to be cost-effective using accepted thresholds. At the 12-month follow-up, the E-couch dominated the waiting list control with more QALYs gained and less costs. Taking into consideration societal costs because of sick leave from work, the E-couch intervention was cost saving at both 6-week and 12-month follow-ups and, therefore, dominated the waiting list control.

Table 3. Health care utilization and other costs at 6 weeks (£).

Group	General practitioner attendance costs, mean (SD)	Outpatient attendance costs, mean (SD)	Inpatient costs, mean (SD)	Cost of work days lost to sick leave, mean (SD)	Mean health care cost (SD)	Mean societal cost (SD)
E-couch (n=383)	38 (89.52)	45.43 (123.99)	72.77 (461.03)	106.65 (476.99)	156.19 (527.42)	264.26 (764.82)
Waiting list (n=761)	38.55 (65.89)	42.49 (111.43)	101.64 (570.51)	123.78 (477.48)	182.67 (643.67)	308.38 (871.52)

Table 4. Health care utilization and other costs at 12 months (£).

Group	General practitioner attendance costs, mean (SD)	Outpatient attendance costs, mean (SD)	Inpatient costs, mean (SD)	Cost of work days lost to sick leave, mean (SD)	Mean health care cost (SD)	Mean societal cost (SD)
E-couch (n=324)	101.10 (152.66)	117.61 (319.44)	207.92 (864.18)	379.47 (1740.74)	425.30 (1077.82)	806.08 (2198.79)
Waiting list (n=680)	106.34 (198.02)	141.84 (416.68)	242.16 (1000.48)	375.18 (1248.74)	490.34 (1264.51)	869.43 (1823.07)

Table 5. Health status and quality-adjusted life years at baseline, 6 weeks and 12 months.

Group	Baseline	6 weeks		12 months	
	SF-6D[a], mean (SD)	SF-6D, mean (SD)	QALY[b], mean (SD)	SF-6D, mean (SD)	QALY, mean (SD)
E-couch	0.60 (0.10)[c]	0.64 (0.12)[d]	0.072 (0.012)[d]	0.66 (0.12)[e]	0.635 (0.10)[e]
Waiting list	0.60 (0.10)[c]	0.62 (0.11)[f]	0.070 (0.011)[f]	0.64 (0.12)[g]	0.621 (0.10)[g]

[a]SF-6D: six-dimensional health state short form.

[b]QALY: quality-adjusted life year.

[c]n=1061.

[d]n=377.

[e]n=324.

[f]n=753.

[g]n=675.

Discussion

Principal Findings

Our findings showed that this fully self-guided internet intervention gave a small reduction in social anxiety symptoms in participants recruited online from the general population, compared with a usual care waiting list control group, and this small but positive finding was robust to the sensitivity analyses, which explored our missing data assumptions. There was a similarly small but significant improvement in fear of negative evaluation. These improvements were also found in the CACE analyses and maintained at the 12-month follow-up. In the context of a very common mental health problem, this finding suggests that automated self-help delivered via the internet could reduce the overall level of social anxiety symptoms in the population, although at an individual level, the mean symptomatic benefit is small (d=0.2). The study findings provide no evidence as to whether this fully self-guided approach has

a role in a clinical setting, where, to date, the evidence base suggests that although unguided self-help has effectiveness, therapist-guided and therapist-led approaches are likely to be superior. The cost-effectiveness analysis demonstrated that the intervention is likely to be cost-effective in both the short and long term, although the gain in general health status and QALY score was very small. The benefit seen in the condition-specific social anxiety outcome measures was greater than the general health status used in the cost utility analysis. Furthermore, the intervention cost could be substantially reduced if the E-couch is used by large numbers at a population level as a public health tool. Given that many people with social anxiety symptoms do not seek help, and that therapist-supported approaches are limited in supply, the findings suggest that unguided digital intervention for social anxiety can be beneficial for some people who do not access professional help and who are increasingly seeking support from apps and other digital tools. The self-help approach tested here might also complement face-to-face therapy, potentially reducing the amount of therapist contact time required and perhaps helping to maintain engagement, although these suggestions need to be empirically tested in future effectiveness and cost-effectiveness work.

Comparison With Prior Work

This study adds to the body of work showing small positive effects for unguided digital self-help for social anxiety [5,9-14] and a range of other mental health problems [29]. Our effect size is smaller than that reported by others. Previous studies have had far fewer participants and usually required them to meet diagnostic criteria for social anxiety disorder. Our aim was to undertake a pragmatic trial addressing social anxiety symptoms (rather than disorder) among individuals in the general population. Our broad inclusion criteria, recruiting volunteers from the general population through internet adverts, including those with symptoms not reaching a diagnostic threshold, are likely to have contributed to the more modest benefit compared with previous work. We made the additional decision to conduct the trial in a fully automated and naturalistic way with no researcher contact to encourage intervention use. Our approach was intended to reflect the real-world situation of members of the public self-selecting digital tools and using them with no contact with health services.

Strengths and Limitations

This study exemplifies both the strengths and weaknesses of undertaking online trials for digital interventions. We were able to recruit large numbers of participants from the general population using digital advertising, and we were able to deliver all measures and the intervention remotely, using little resources and with no requirement for any "real-world" contact between participants and researchers. The flipside of this was that, in common with other fully automated trials of unguided online interventions, there were high levels of dropout from the intervention and attrition from the trial [30]. This is commonly seen in internet research [31], including the higher level of retention in the control arm [32], which may be partly explained by these participants being on a waiting list and, therefore, having an incentive to keep returning to complete measures, and partly by participants in the intervention arm being required to "take action" (work through the intervention), whereas the control group could be more passive. Other possible reasons for dropout include some participants not liking the intervention, or feeling it was not working, or indeed dropping out because they felt they had improved and no longer needed it. The high loss to follow-up was compounded at the 3- and 6-month follow-up points by a software glitch, which reduced the number of emails sent to participants at this time. Fortunately, these were always intended as interim time points measured to contribute to the overall mixed linear model. We undertook sensitivity analyses and explored various approaches to adjusting for the missing data. All outcome measures were self-report with no observer-rated objective assessment. This was in line with our desire to deliver a fully automated trial, and the scales are well validated, but the subjective nature of these measures is a potential source of response bias. We did not employ a placebo but instead used a waiting list comparator whereby people received "usual care." In other work, educational website placebos have often demonstrated an active effect [33]. Our pragmatic choice of control group, given that participants were not blind to allocation, may have introduced bias and increased the likelihood of a beneficial effect. Finally, most participants in this study were women. Social anxiety symptoms are twice as common in women than men [4], and women are more likely to seek health care generally [34]. Further work on the predictors and mediators of both adherence and response would be valuable [35,36].

Conclusions

For people with social anxiety symptoms in the general population who are not receiving other forms of help, an online unguided tool based on cognitive behavioral principles accessed via a computer or mobile phone gave a small but significant improvement in social anxiety symptoms compared with no intervention. As with many online trials of digital interventions, we experienced a "methodological trade-off" between having a cheap, scalable model of intervention delivery versus the statistical challenge of a high degree of missing data. Our findings suggest this intervention could potentially offer the first self-help rung on the ladder of a stepped approach to social anxiety symptoms.

Acknowledgments

The authors would like to thank the independent members of our Trial Steering Committee: Dr Peter Davidson MBBS, MSc (Chair), formerly Director of the National Institute for Health Research Evaluation, Trials and Studies Coordinating Centre, Chilworth, Southampton, UK; Dr Ceire Costelloe, PhD, Director of Global Digital Health Unit, Imperial College London, UK; Dr Adam Geraghty, PhD, Faculty of Medicine, University of Southampton, Southampton, UK; and Mr Gareth Stephens, lay representative. They would also like to thank Mr Anthony Bennett, the Chief Technology Officer at ehub Health, part of the team

who developed and implemented the trial software, for technical expertise and support. The authors would also like to thank the participants who took part in this study.

This work was funded by the MQ: Transforming Mental Health charity under its PsyIMPACT funding call (MQ14PE_25). JP is also funded by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust. YY is supported by the National Institute for Health Research Oxford Biomedical Research Centre. KG was supported by a National Health & Medical Research Council Senior Research Fellowship.

The funding bodies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Authors' Contributions

JP had full access to all study data and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors were responsible for study concept; design, acquisition, analysis, or interpretation of data; critical revision of the manuscript for important intellectual content; and administrative, technical, or material support. JP was responsible for drafting of the manuscript. JM, MS, and LY were responsible for statistical analysis. YY was responsible for economic evaluation. JP and KG were responsible for obtaining funding. JP and KG were responsible for study supervision.

Conflicts of Interest

All authors have completed and submitted the International Committee of Medical Journal Editors Form for Disclosure of Potential Conflicts of Interest. ehub Health Pty Ltd, a spinout from the Australian National University (ANU) has been granted the license to the E-couch intervention with royalties returning to the ANU. KB is an owner and employee of ehub Health Pty Ltd. As a cocreator of E-couch, KG is entitled to a percentage of any royalties received by the ANU from ehub Health Pty Ltd. She has no other financial interest in ehub Health Pty Ltd but is an honorary scientific advisor to the company. JP is guarantor of this paper and has no interests to declare in the E-couch intervention and KB and KG were not involved in the statistical analysis. No other disclosures were reported.

Multimedia Appendix 1
Study protocol.
[DOCX File , 733 KB - jmir_v22i1e16804_app1.docx]

Multimedia Appendix 2
Sensitivity, subgroup and additional analyses.
[DOCX File , 256 KB - jmir_v22i1e16804_app2.docx]

Multimedia Appendix 3
Descriptive statistics of Social Phobia Inventory (SPIN-17) score at all time points by randomized group.
[DOCX File , 18 KB - jmir_v22i1e16804_app3.docx]

Multimedia Appendix 4
CONSORT ‐ EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 394 KB - jmir_v22i1e16804_app4.pdf]

References

1. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676 [FREE Full text] [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
2. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Sep;17(3):327-335 [FREE Full text] [Medline: 26487813]
3. Crisan LG, Vulturar R, Miclea M, Miu AC. Reactivity to social stress in subclinical social anxiety: emotional experience, cognitive appraisals, behavior, and physiology. Front Psychiatry 2016;7:5 [FREE Full text] [doi: 10.3389/fpsyt.2016.00005] [Medline: 26858658]
4. Fehm L, Beesdo K, Jacobi F, Fiedler A. Social anxiety disorder above and below the diagnostic threshold: prevalence, comorbidity and impairment in the general population. Soc Psychiatry Psychiatr Epidemiol 2008 Apr;43(4):257-265. [doi: 10.1007/s00127-007-0299-4] [Medline: 18084686]
5. Mayo-Wilson E, Dias S, Mavranezouli I, Kew K, Clark DM, Ades AE, et al. Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. Lancet Psychiatry 2014;1(5):368-376. [doi: 10.1016/s2215-0366(14)70329-3]

6. Cuijpers P, van Straten A. Improving outcomes in social anxiety disorder. Lancet Psychiatry 2014 Oct;1(5):324-326 [FREE Full text] [doi: 10.1016/S2215-0366(14)70351-7] [Medline: 26360985]
7. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):603-613. [doi: 10.1001/archpsyc.62.6.603] [Medline: 15939838]
8. Boettcher J, Carlbring P, Renneberg B, Berger T. Internet-based interventions for social anxiety disorder - an overview. Verhaltenstherapie 2013;23(3):160-168. [doi: 10.1159/000354747]
9. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
10. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for Social Anxiety Disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
11. Gershkovich M, Herbert JD, Forman EM, Schumacher LM, Fischer LE. Internet-delivered acceptance-based cognitive-behavioral intervention for social anxiety disorder with and without therapist support: a randomized trial. Behav Modif 2017 Sep;41(5):583-608. [doi: 10.1177/0145445517694457] [Medline: 28776431]
12. 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]
13. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a web-based social anxiety intervention among university students: randomized controlled trial. J Med Internet Res 2018 Mar 21;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
14. Kählke F, Berger T, Schulz A, Baumeister H, Berking M, Auerbach RP, et al. Efficacy of an unguided internet-based self-help intervention for social anxiety disorder in university students: a randomized controlled trial. Int J Methods Psychiatr Res 2019 Jun;28(2):e1766. [doi: 10.1002/mpr.1766] [Medline: 30687986]
15. Bowler JO, Mackintosh B, Dunn BD, Mathews A, Dalgleish T, Hoppitt L. A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: effects on anxiety, depression, attentional control, and interpretive bias. J Consult Clin Psychol 2012 Dec;80(6):1021-1033 [FREE Full text] [doi: 10.1037/a0029932] [Medline: 22963595]
16. Connor KM, Davidson JRT, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. Br J Psychiatry 2000 Apr;176:379-386. [doi: 10.1192/bjp.176.4.379] [Medline: 10827888]
17. Carleton RN, Collimore KC, McCabe RE, Antony MM. Addressing revisions to the Brief Fear of Negative Evaluation scale: measuring fear of negative evaluation across anxiety and mood disorders. J Anxiety Disord 2011 Aug;25(6):822-828. [doi: 10.1016/j.janxdis.2011.04.002] [Medline: 21565463]
18. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
19. Fat LN, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Qual Life Res 2017 May;26(5):1129-1144 [FREE Full text] [doi: 10.1007/s11136-016-1454-8] [Medline: 27853963]
20. Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): group and individual level analysis. Health Qual Life Outcomes 2012 Dec 27;10:156 [FREE Full text] [doi: 10.1186/1477-7525-10-156] [Medline: 23270465]
21. Ware JE. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. New York: Routledge; 1999.
22. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
23. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91(434):444-455. [doi: 10.1080/01621459.1996.10476902]
24. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clin Trials 2007;4(2):125-139. [doi: 10.1177/1740774507077849] [Medline: 17456512]
25. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. PLoS One 2011 Mar 9;6(3):e14740 [FREE Full text] [doi: 10.1371/journal.pone.0014740] [Medline: 21408060]
26. StataCorp. College Station, Texas: StataCorp LLC; 2017. Stata Statistical Software: Release 15 URL: https://plattogvibinta.cf/stata-statistical-software-release-15.html
27. Curtis L, Burns A. Unit Costs of Health & Social Care 2017. University of Kent, Canterbury, UK: Personal Social Services Research Unit (PSSRU); 2017.

28. Sheffield Licensing - University of Sheffield. Sheffield, UK: University of Sheffield SF-6D: Calculating QALYs from the SF-36 and SF-12. URL: https://licensing.sheffield.ac.uk/i/health-outcomes/SF-6D.html [accessed 2019-10-01]
29. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
30. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
31. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 1;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]
32. 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]
33. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
34. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract 2016 Mar 31;17:38 [FREE Full text] [doi: 10.1186/s12875-016-0440-0] [Medline: 27036116]
35. Boettcher J, Renneberg B, Berger T. Patient expectations in internet-based self-help for social anxiety. Cogn Behav Ther 2013;42(3):203-214. [doi: 10.1080/16506073.2012.759615] [Medline: 23697570]
36. Nordgreen T, Havik OE, Ost LG, Furmark T, Carlbring P, Andersson G. Outcome predictors in guided and unguided self-help for social anxiety disorder. Behav Res Ther 2012 Jan;50(1):13-21. [doi: 10.1016/j.brat.2011.10.009] [Medline: 22134140]

Abbreviations

ANU: Australian National University
BFNE-S: Brief Fear of Negative Evaluation
CACE: Complier Average Causal Effect
NHS: National Health Service
QALY: quality-adjusted life year
SPIN-17: 17-item Social Phobia Inventory
SWEMWBS: Short Warwick-Edinburgh Mental Well-Being Scale

Edited by G Eysenbach; submitted 26.10.19; peer-reviewed by N Van-Zalk, C Bedard; comments to author 16.11.19; revised version received 22.11.19; accepted 10.12.19; published 10.01.20.

Please cite as:
Powell J, Williams V, Atherton H, Bennett K, Yang Y, Davoudianfar M, Hellsing A, Martin A, Mollison J, Shanyinde M, Yu LM, Griffiths KM
Effectiveness and Cost-Effectiveness of a Self-Guided Internet Intervention for Social Anxiety Symptoms in a General Population Sample: Randomized Controlled Trial
J Med Internet Res 2020;22(1):e16804
URL: http://www.jmir.org/2020/1/e16804/
doi:10.2196/16804
PMID:31821151

Original Paper

A Web- and Mobile App–Based Mental Health Promotion Intervention Comparing Email, Short Message Service, and Videoconferencing Support for a Healthy Cohort: Randomized Comparative Study

Melanie Elise Renfrew[1], BEd, Grad Dip; Darren Peter Morton[1], MAppSci, PhD, FASLM; Jason Kyle Morton[1], BEd, BSci, PhD; Jason Scott Hinze[1], BEd, MLMEd, PhD; Peter James Beamish[1], BEd, MEdSt, PhD; Geraldine Przybylko[1], BCom, MBA, MPH; Bevan Adrian Craig[1], BN, Grad Dip, PhD

Lifestyle and Health Research Centre, Avondale University College, Cooranbong, New South Wales, Australia

Corresponding Author:
Melanie Elise Renfrew, BEd, Grad Dip
Lifestyle and Health Research Centre
Avondale University College
582 Freemans Drive
Cooranbong, New South Wales,
Australia
Phone: 61 405445151
Email: melrenfrew@gmail.com

Abstract

Background: The rapid increase in mental health disorders has prompted a call for greater focus on mental health promotion and primary prevention. Web- and mobile app–based interventions present a scalable opportunity. Little is known about the influence of human support on the outcomes of these interventions.

Objective: This study aimed to compare the influence of 3 modes of human support on the outcomes (ie, mental health, vitality, depression, anxiety, stress, life satisfaction, and flourishing) of a 10-week, Web- and mobile app–based, lifestyle-focused mental health promotion intervention among a healthy adult cohort.

Methods: Participants were recruited voluntarily using a combination of online and offline advertising. They were randomized, unblinded into 3 groups differentiated by human support mode: Group 1 (n=201): standard—fully automated emails (S); Group 2 (n=202): standard plus personalized SMS (S+pSMS); and Group 3 (n=202): standard plus weekly videoconferencing support (S+VCS), hosted by 1 trained facilitator. Participants accessed the intervention, including the questionnaire, on a Web-based learning management system or through a mobile app. The questionnaire, administered at pre- and postintervention, contained self-reported measures of mental well-being, including the "mental health" and "vitality" subscales from the Short Form Health Survey-36, Depression Anxiety and Stress Scale-21, Diener Satisfaction With Life Scale (SWLS), and Diener Flourishing Scale.

Results: Of 605 potential participants, 458 (S: n=157, S+pSMS: n=163, and S+VCS: n=138) entered the study by completing registration and the preintervention questionnaire. At post intervention, 320 out of 458 participants (69.9%; S: n=103, S+pSMS: n=114, and S+VCS: n=103) completed the questionnaire. Significant within-group improvements were recorded from pre- to postintervention in all groups and in every outcome measure ($P \leq .001$). No significant between-group differences were observed for outcomes in any measure: mental health ($P=.77$), vitality ($P=.65$), depression ($P=.93$), anxiety ($P=.25$), stress ($P.57$), SWLS ($P=.65$), and Flourishing Scale ($P=.99$). Adherence was not significantly different between groups for mean videos watched ($P=.42$) and practical activity engagement ($P=.71$). Participation in videoconference support sessions (VCSSs) was low; 37 out of 103 (35.9%) participants did not attend any VCSSs, and only 19 out of 103 (18.4%) attended 7 or more out of 10 sessions. Stratification within the S+VCS group revealed that those who attended 7 or more VCSSs experienced significantly greater improvements in the domains of mental health ($P=.006$; $d=0.71$), vitality ($P=.005$; $d=0.73$), depression ($P=.04$; $d=0.54$), and life satisfaction ($P=.046$; $d=0.50$) compared with participants who attended less than 7.

Conclusions: A Web- and mobile app–based mental health promotion intervention enhanced domains of mental well-being among a healthy cohort, irrespective of human support. Low attendance at VCSSs hindered the ability to make meaningful between-group comparisons. Supplementing the intervention with VCSSs might improve outcomes when attendance is optimized.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR): 12619001009101; http://www.anzctr.org.au/ACTRN12619001009101.aspx

(J Med Internet Res 2020;22(1):e15592) doi:10.2196/15592

KEYWORDS

guidance; health promotion; eHealth; short message service; videoconferencing

Introduction

Background

In 2012, the World Health Assembly requested the development of a plan to address escalating mental distress, resulting in the Mental Health Action Plan 2013-2020, which asserts that there is "no health without mental health" [1]. However, as we approach 2020, mental well-being continues to deteriorate. Depression is the leading cause of disability worldwide [2,3], and its prevalence is rising [4], evidenced by increasing antidepressant use [5] and the high rate of suicide—the second leading cause of death in the 15 to 29 years age group internationally [3]. Comorbidity is common [6], and those with major depression have a 40% to 60% risk of dying prematurely [1]. A 2018 review [7] of 12 countries revealed that depression and anxiety are the third most common reason for visits to a general practitioner in the developed world.

Mental well-being, even in the seemingly healthy population, is compromised by mounting stress and anxiety. In the 2015 Australian Psychological Society's stress and well-being survey [8], 35% of those surveyed reported significant distress levels, 26% reported above normal anxiety, and 26% described themselves as having moderate to extreme depression levels. Therefore, provision of mental health promotion interventions (MHPIs) may be a vital approach to enhance mental well-being among healthy (ie, nonclinical), community-based cohorts.

There is a growing evidence base supporting the efficacy of MHPIs to improve the mental well-being of nonclinical population groups. These include interventions to alleviate stress [9-11], mindfulness training [12-15], lifestyle medicine strategies such as nutrition [16,17], and exercise [18,19] and an array of positive psychology interventions [20-22]. A recent analytic review [23] highlighted strategies from the fields of positive psychology and lifestyle medicine that have demonstrated effectiveness in enhancing mental well-being. Combining strategies from both fields, the same author devised a lifestyle-focused, multimodal intervention that has been piloted among university students on 2 occasions [24,25] and more recently as a Web-based randomized controlled trial (RCT) for a healthy community cohort [26]. The Web-based RCT demonstrated significantly greater improvements (P<.001) than the control group for all outcomes, with medium effect size improvements for the intervention arm in mental health (d=0.52), vitality (d=0.56), and stress (d=0.45). The same intervention was used for this study.

Web- and mobile app–based technology offers an unprecedented opportunity for disseminating MHPIs to healthy cohorts in community settings. Advantages, when compared with face-to-face interventions, include cost-effectiveness [27,28], accessibility, and scalability [29,30]. Furthermore, up to 80% of potential users are "e-preferers" [31]. However, digital delivery also poses unique challenges (eg, quality control, data control, high dropout attrition, and low adherence) that require creative solutions [32,33]. Notably, provision of human support (ie, guidance) is recognized as a possible modulating influence on adherence and outcomes [33,34].

Human Support

Human support provision in Web-based interventions has generally being associated with improved adherence and outcomes in clinical cohorts [32,35-39]. A meta-analysis of 12 studies treating depression found effect sizes for studies that included human support were larger (d=0.61) in comparison with unsupported studies (d=0.25) [38]. Nevertheless, some studies have found that the level of human support does not significantly influence outcomes [40-44]. Furthermore, variability in the provision of human support (eg, mode, intensity, and synchronicity) results in high heterogeneity, which makes comparisons between studies problematic [33,39].

It is imperative to investigate the role of human support among nonclinical groups. Support requirements may be markedly different among healthy cohorts. Symptomatic factors (eg, apathy, weakened motivation, and general malaise) that could impede a clinical cohort from completing a program may be nonexistent or differ considerably for a nonclinical group. Conceivably, MHPIs may prove to be a credible pathway to enhance mental well-being and serve as a vital buffer to protect healthy populations against mental distress [45].

Studies comparing human support factors for healthy cohorts engaging in Web- or mobile app–based mental well-being interventions are scarce, and outcomes are mixed. Several primary prevention studies (all classified as "indicated" prevention) have reported no significant difference between supported and unsupported arms [46-48]. Zarski et al [49] examined 3 support approaches and compared the effects on adherence with a stress management intervention. Monitoring and written feedback improved adherence. However, administrative support (ie, access to a support team for technical assistance) had no positive effect [49]. Allexandre et al [50] compared no support, group support, and group support with added expert clinical support for a stress management intervention. Group support was beneficial, but added clinical care contributed no extra benefit. However, the program content was Web-based, and support was provided face-to-face in a work setting, making comparison problematic. Finally, a review and metanalysis of 23 Web- and computer-based interventions to alleviate stress found that supported interventions demonstrated greater effects on outcomes (d=0.64) than unsupported (d=0.33) [51].

A confounder when drawing conclusions about the impact of human support on the effectiveness of Web- and computer-based programs is that there are many modes and delivery styles classified as human support, which differ markedly in their resource requirements (eg, time, cost, and intensity). For this study, we chose 3 support modes to compare, based on low (automated emails), medium (personalized SMS messaging), and high (facilitated videoconference) resource requirements.

Email Support

Emails are widely used [33] and easily incorporated into a Web-based intervention as an asynchronous, low intensity, low cost support method [52]. Notwithstanding, heterogeneity in the way emails are utilized makes comparison difficult, and results are mixed in clinical settings [46,53-55]. For instance, automated emails, often used as engagement prompts, may be built into the system design and require virtually no monitoring once set up. Conversely, personally tailored, individually written emails require effort and time on behalf of the support person and may be considered a more intense mode of support [52].

Short Message Service Support

SMS support varies in intensity depending on the method of dissemination (ie, automated or individualized), however, it is easily accessible and portable [56] through the widespread use of mobile phones. Researchers have used SMS support to aid adherence to medication [57,58], support asthma treatment [59], and promote adherence to healthy lifestyle practices [60] among other uses. A systematic review and meta-analysis of Web-based interventions, that used additional support modes demonstrated that SMS had large effects (d=0.81) compared with phone (d=0.35) and email (d=0.18) [61]. Nevertheless, users may ignore SMS prompts when they are perceived as impersonal, too frequent, or automated [62].

Videoconferencing Support

Videoconferencing most closely replicates the face-to-face setting; however, it requires a greater investment of time, cost, and human resources. Notwithstanding, it may provide a feasible, personal, and acceptable mode of support, similar to face-to-face settings, as long as technical assistance is available [63]. Participants value ease of accessibility and are still able to bond as a group despite lack of various communication cues (eg, body language) [63]. Videoconferencing has been successfully used to support caregivers [64], patients with chronic disease [65], new parents [66], and bariatric surgery patients [67], among others.

Identifying the optimum input of human support resources to maximize program effectiveness is an important consideration for researchers designing MHPIs for nonclinical cohorts. This comparative study seeks to add to the evidence base by asking the question: "What is the influence of different modes of human support on the outcomes of a Web- and mobile app–based MHPI for a healthy, community-based cohort?"

Methods

Recruitment

Participants were recruited from an Australian and New Zealand faith-based cohort. Advertising, conducted from July to September 2018, included offline marketing through periodicals, magazines, and bulletins of the faith-based organization and online marketing using a combination of website, social media, and email strategies. Advertising material directed potential applicants to a Web page to examine the inclusion criteria (Textbox 1), acquaint themselves with participation expectations, and fill out an enrollment application.

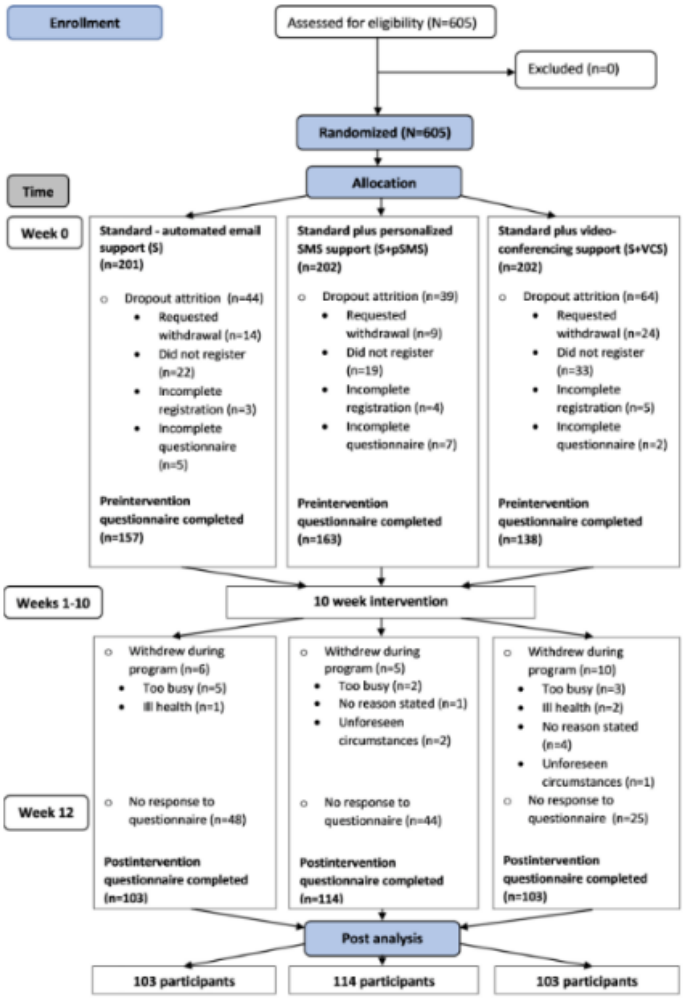
Approved applicants were randomized into 1 of 3 intervention groups (Figure 1) using computer-generating software. An email notified applicants of acceptance into the study, group allocation, instructions detailing the steps to complete registration on the Web-based electronic learning management system (eLMS), and a unique link to activate the registration process. Informed consent was gained as part of the registration procedure. Once registered, users could opt to download the mobile app and access all features of the intervention, including self-reported questionnaires, from either the eLMS or from the mobile app.

To provide anonymity, participants were permitted to use a pseudo name for the duration of the intervention, if desired. In the S+VCS group, which involved videoconferencing using the app "Zoom," the participants were given the option to choose a pseudo name and have the camera switched off.

Textbox 1. Inclusion criteria.

- Aged 18 years or older
- Mobile phone with SMS capability
- Internet access
- Australian or New Zealand resident
- Fluent in English
- Acceptance to provide informed consent
- Permission given for anonymous data to be used for research

Figure 1. Flow of participants.



Study Design

The study was a multiarm, randomized comparative design with 3 intervention groups that differed according to modes of human support:

1. Group 1: Standard—fully automated emails (S).
2. Group 2: Standard plus personalized SMS support (S+pSMS).
3. Group 3: Standard plus videoconferencing support (S+VCS).

Figure 1 shows that participants in each group accessed the intervention simultaneously, from September to December 2018.

Intervention

The 10-week intervention is presented in an audio-visual format and was initially administered in 2017 as a Web-based program on an eLMS. Version 2, released immediately before commencement of this study (August 2018), included an updated eLMS and a mobile app called "MyWellness." See Multimedia Appendix 1 for screenshots. No new features were introduced during the study period.

The lifestyle-focused intervention is underpinned by the theory of planned behavior (TPB), which has demonstrated substantial health behavior change benefits [61]. The intervention actively seeks to change behavior through education and increase perceived control through easily attainable weekly challenges [24]. Using the pedagogical framework of Learn, Experience, Think, Share, the participants were introduced to basic neuroscience and explored 10 evidence-based strategies from lifestyle medicine and positive psychology for improving mental well-being. Table 1 outlines the weekly topics covered throughout the intervention [68-91].

Participants accessed the intervention free of charge by logging onto the eLMS or the mobile app. Each participant's personal dashboard provided access to the weekly session, which included approximately 30 min of audio-visual content, a downloadable workbook, and extra reading materials relating to the topic [92]. The intervention emphasized experiential learning by encouraging participants to engage in daily and weekly challenges through which they could log their activity to earn challenge points. Participants were emailed a "timetable overview" (Multimedia Appendix 2), which outlined the recommended schedule to complete the intervention within the 10-week timeframe.

Table 1. Intervention overview.

Week	Topic	Overview	Daily challenge	Weekly challenge
1	Speak positively (your limbo is listening)	• Limbic system introduction—the "emotional brain" • Limbic system is "wired" to language area of brain [68,69]	Offer a genuine compliment.	Memorize an inspirational text or saying.
2	Move dynamically (motion creates emotion)	• Proprioceptors (nerve cells that detect movement) pass through the limbic system • Movement (even just 10 min) improves mood [70-73]	Complete 30 min of moderate-intensity activity.	Perform resistance exercises once during week (exercises demonstrated on video).
3	Immerse in an uplifting natural environment (blue and green should often be seen)	• The limbic system receives messages from all the senses • The limbic system likes blue and green spaces (ie, natural settings) • The limbic system needs about 30 min of 10,000 LUX of light daily [74-76]	Immerse in an uplifting natural environment for 30 min daily.	Watch a sunrise from an appealing blue or green location.
4	Immerse in a positive social environment (together feels better)	• Limbic systems communicate • Create positive social environments by making new friends or strengthening existing relationships • Forgiveness [77-79]	Daily, do something intentional to show a friend or family member they are loved (use their love language).	Give up your right to hurt someone who has hurt you (Act of forgiveness).
5	Look to the positive (feelings follow your focus)	• "Emotional brain" is wired to "Thinking brain" • What you focus on affects how you feel • Upward or downward spirals [21,45,80-82]	Write down 3 things that went well today "What Went Well?"	Gratitude visit—identify someone you are grateful to, write a gratitude letter to them, deliver and read in person, if possible.
6	Eat nutritiously (food feeds your mood)	• Gut bacteria linked to mood • Feed gut bacteria high fiber diet • Plant-based foods are high fiber • Eat a wide variety of fruit, vegetables, legumes, and grains [16,17,83,84]	Eat 8 fists full of fiber daily.	Prepare and share a high-fiber plant-based meal with one or more friends.
7	Rest—sleep (rest to feel your best)	• Sleep is fundamental for feeling upbeat (7-8 hours optimal) • Blue light vs yellow/orange light • Caffeine, lack of physical activity, and blue light—deprived sleep [85,86]	Spend 8 hours in bed every night.	Spend an evening by firelight.
8	Rest—from stress (stress less)	• SMILERS strategies "open the valves" • Physical activity, practicing mindfulness, laughing, and rest day [13,51,87,88]	15-min "sit in silence" mindful activity.	Take a "guilt-free" day off.
9	Serve others (giving is living)	• Contributing/serving is emotionally uplifting • Serve sustainably using signature strengths [89-91]	Perform 1 or more random acts of kindness each day.	Use your significant strength to perform a significant act of service.
10	What does it take to flourish?	Five areas to flourish: PEARM—positive emotions, engagement, achievement, relationships, and meaning [90]	Spend time engaging in something you enjoy.	Create a list of goals and an action plan to achieve them.

Persuasive Systems Design Features

Features of Persuasive Systems Design (PSD) were key to the intervention, and numerous support principles were utilized on the Web platform and mobile app to encourage behavior change. The PSD model, developed by Oinas-Kukkonen and Harjumaa [93], is largely based on earlier work by Fogg [94]. The model outlines a taxonomy of 28 principles, classified under 4 key areas of support, for designing completely digital, behavioral change interventions. "Primary Task Support" components of PSD included the use of "reduction" to simplify key learnings into practical strategies for implementation, along with "self-monitoring" and "rehearsing" desired behaviors in the form of daily and weekly challenges that were recorded on the eLMS or app. An element of "Dialogue Support" [93] utilized was "rewards" in the form of badge icons. "Social Support" elements of PSD [93] provided opportunities for various forms of interaction. Participants could choose to post challenge activity to a public "feed" (viewable by other group members) providing "social learning" and "social facilitation" opportunities. Furthermore, challenges posted as "public" provided opportunity for interaction between group members. "Competition" and "social comparison" were aroused through a challenge points leader board.

Human Support

Human support was crafted to reflect values of the Supportive Accountability Model (SAM), developed by Mohr et al [34]. They propose a range of support measures to improve adherence, including focusing on process expectations rather than outcome expectations, positive feedback provision, avoidance of controlling behavior, timely communication, being trustworthy, facilitating engaging discussions, and avoidance of pressure tactics [34].

Advertising Period

During the advertising period, the mobile app "Hitemup" was used to send enrollees (ie, those who had registered interest in joining the study) weekly, personalized SMS communication to remind them of the upcoming program.

Orientation and Registration Period

During orientation week, all enrollees were sent an email link to register onto the eLMS to finalize registration and complete the preintervention questionnaire. A total of 2 SMS reminders were sent, several days apart, as reminders to register.

On the intervention start date, a "welcome" email and SMS were sent to all registered participants outlining the steps to access the first session. In addition, within the first 2 weeks of the intervention, those who had not yet completed registration were sent 4 reminder SMSs and email messages inviting them to access the provided link to complete registration. Once participants completed registration, human support reverted to the constraints outlined for each group, and the eLMS was used to send SMS and automated email messages to registered participants.

Group 1 Support: Automated Emails—Standard

Automated emails, which are the routine communication mode outside the research setting, were set up as part of the eLMS. All participants received a weekly email on the day before the next session commencing. The email included a link to a 20 to 25 second video by the presenter, inviting them to engage with the next presentation. A total of 3 days after a lesson was released, the system checked to see if the participant had logged any challenges: if "yes," participants were sent an email validating their participation, if "no," an email prompted the participant to complete the relevant challenge. A total of 8 days after a new lesson was released, participants who had not watched the video were sent an email reminder.

Group 2 Support: Standard Plus Personalized SMS Messages)

The S+pSMS group received automated emails plus SMS messages that were sent 3 times weekly for the first 3 weeks and then 2 times weekly for the remaining 7 weeks. The reduction from 3 messages per week to 2 messages per week recognized the notion that support may have a threshold [34]; too many messages may be seen as controlling and undermine the commitment of individuals, therefore impairing engagement [34]. The messages focused on process accountability (ie, completing target behaviors such as viewing content and engaging in experiential learning) as opposed to outcome accountability, which may be detrimental and beyond an individual's control [34].

Each SMS included the participant's first name, a predetermined message to prompt engagement, and was signed by the research team member (MR) who also provided technical assistance. Specific message content included different combinations of the following: a brief hook sentence (eg, Lesson 2—Learn 3 practical tips to calm your nerves, Lesson 3—Have you had your dose of lux today?), the link to log into the eLMS, an inspiring quote, tips for implementing challenge activity, and a phone number for technical assistance.

Group 3 Support: Standard Plus Videoconference Support

As well as automated email support, participants in the S+VCS group were invited to attend a synchronous, videoconference session using the app "Zoom." A weekly timetable (Multimedia Appendix 3) provided 9 different time possibilities, and participants were invited to attend any one of these. During orientation week, participants were encouraged to attend 1 of 9 available "tech check" sessions. Each week, participants were sent 1 reminder SMS and an email with a link to the online meeting. Discussion sessions lasted 20 to 30 min and were led by an experienced educator who holds a post-graduate degree in lifestyle medicine. The facilitator had previously hosted many preventative health videoconferences for participant groups doing the renowned Complete Health Improvement Program. Prior training included 6 mentoring sessions on how to facilitate successful videoconferences in a Web-based setting. The videoconference support sessions included a recap of the weekly content by the facilitator, sharing new learnings and challenge experiences, plus dialogue on ways to incorporate strategies learned into everyday life.

Measurement—Outcomes

Participants completed a self-report questionnaire, the "7 Dimensions of Wellness Index" (7DWI), at preintervention (week 0) and postintervention (week 12). The 7DWI combines demographic- and lifestyle-related questions with various, freely available, validated instruments that measure 7 well-being domains (emotional, physical, social, spiritual, vocational, intellectual, and environmental). For the purpose of this study, the following instruments were utilized from the 7DWI to measure aspects of mental well-being:

36-Item Short Form Survey

The 36-item Short Form Health Survey (SF-36) is suitable for use in general population surveys [95], and 2 subscales of the SF-36 were used for this study: mental health (5 items) and vitality (4 items) [96]. Good internal consistency and reliability across the whole survey and subscales have been reported [97]. Cronbach alpha scores of .90 for the mental health subscale and .87 for the vitality subscale are well above the minimum reliability standard (.50-.70) [98].

21-Item Depression Anxiety and Stress Scale

The 21-item Depression Anxiety and Stress Scale (DASS-21) is widely used to measure depression, anxiety, and stress (7 items per factor), and satisfactory reliability has been demonstrated for both clinical and nonclinical samples [99-101]. The DASS has demonstrated adequate internal consistency as a total scale (Cronbach alpha >.90) [102] and for the 3 subscales, with Cronbach alpha scores ranging from .76 to .91 [99,101].

Satisfaction With Life Scale

Diener 5-item Satisfaction With Life Scale (SWLS) measures global life satisfaction and was initially tested on a nonclinical group of university students [103]. Good internal consistency was found (Cronbach alpha .87) [103], and the SWLS was deemed suitable for a wide range of age groups and settings [104]. Data from 6 studies indicated that the scale has high internal consistency (Cronbach alpha ranged from .79-.89) [105]. Furthermore, a meta-analysis of 62 articles provided 76 reliability coefficients testing the SWLS. The mean Cronbach alpha was .78—signifying good internal consistency [106].

Flourishing Scale (Diener)

Diener brief 8-item Flourishing Scale, designed to measure "social-psychological prosperity," is suitable for nonclinical cohorts, correlates highly with other measures of well-being, and demonstrates good internal consistency under initial analysis (Cronbach alpha .87) [107]. The Flourishing Scale has a unidimensional factor structure, and more recent analysis, across 2 samples, also showed good consistency (.78 and .83) [108].

Measurement—Adherence

Primary adherence was measured as the total number of weekly videos viewed out of a possible ten presentations. Challenge adherence was also measured, with participants able to accumulate points through practical daily and weekly challenge activities. Each daily challenge was worth 10 points, and weekly challenges were worth 30 points. Participants could score a maximum of 100 points each week throughout the 10-week intervention, thereby accumulating a total of 1000 points to be considered fully adherent. In addition, for the S+VCS group, videoconference attendance was recorded as a score out of 10.

Sample Size

The a priori power analysis indicated a required sample size of 148 participants in each arm of the study. This calculation was based on the following assumptions or requirements and relied on data from previous studies [24,26]: (1) participants were to be allocated equally between the 3 arms; (2) an ability to detect a 15% improvement in the mental health and vitality scales, as this was considered a clinically significant outcome [24]; (3) a 30% attrition rate, based on attrition rates observed in a prior similar study [26]; (4) a power level of ≥80%; (5) significance level of 0.05 (95% CI); and (6) a distribution SD of 16.1 in the mental health subscale, based on a prior study [24].

Randomization

Participants were randomized by a person not on the research team using computer random number generation. Equitable distribution of age and gender was tested and confirmed, negating the need for further stratification. The 3 randomized groups were allocated to a study arm by the person who conducted randomization. Researchers and participants were unblinded, and participants were notified of the mode of support they would receive during registration. As there was no control group, all groups were comparators of interest.

Statistical Analysis

Data were analyzed using the IBM SPSS Statistics Software (version 25). Within-group changes from baseline to post intervention were calculated using paired *t* tests, and repeated measures generalized linear models were used to measure analysis of variance between groups, taking into account time and group effects. Cohen *d* was used to measure effect size, and Fisher Exact test was used to test for relationships between the categorical variables. The 458 participants who completed the preintervention questionnaire were included in the analysis of baseline characteristics. All remaining analyses included only the participants who completed both preintervention and postintervention measures (n=320). This manuscript was prepared according to Consolidated Standards of Reporting Trials (CONSORT) guidelines [109] and utilized the CONSORT-EHEALTH checklist (Multimedia Appendix 4).

Ethics and Informed Consent

Ethics approval was granted from the Avondale Human Research Ethics Committee (Approval No. 2018.09). Prospective participants were emailed an "information statement" outlining the details of the study and an "informed consent statement" notifying them that choosing to register onto the eLMS would signify informed consent.

Results

Participants

Figure 1 records the flow of participants throughout the intervention period. Potential participants (n=605) were enrolled through the information Web page and were randomized into

3 groups: S (n=202), S+pSMS (n=202), and S+VCS (n=201). A total of 458 participants registered on the eLMS, completed the preintervention questionnaire, and were analyzed for baseline characteristics (S=157, S+pSMS=163, and S+VCS=138). At 12 weeks, 320 out of 458 (69.9%) participants had completed both the pre- and postintervention questionnaire required for postanalysis (S=103, S+pSMS=114, and S+VCS=103).

Baseline Characteristics

Table 2 describes the demographic and mental health characteristics of the study group at baseline. Fisher Exact tests demonstrated no significant between-group differences in any of the categorical characteristics of the study population. With a mean age of 45.5 (SD 13.7) years, participants were predominantly female (77.7%, 356/458), white (76.9%, 352/458), and had a tertiary education (78.0%, 357/458). There were no significant differences between the groups in any of the psychometric measures at baseline (Table 2).

Table 2. Baseline characteristics of the study population.

Factor	Standard automated email group (S; n=157)	Standard plus SMS message group (S+pSMS; n=163)	Standard plus videoconference support group (S+VCS; n=138)	Combined groups (N=458)	Analysis of variance or Fisher Exact test, *P* value
Age (years), mean (SD)	45.6 (13.9)	46.5 (13.4)	44.3 (13.9)	45.5 (13.7)	.19
Gender, n (%)					.82
Female	124 (79.0)	126 (77.3)	106 (76.8)	356 (77.7)	
Male	33 (21.0)	37 (22.7)	31 (22.5)	101 (22.1)	
Other	0 (0)	0 (0)	1 (0.7)	1 (0.2)	
Education, n (%)					.16
Primary or elementary	0 (0)	2 (1.2)	0 (0)	2 (0.4)	
Secondary or high school	36 (22.9)	39 (23.9)	24 (17.4)	99 (21.6)	
Tertiary undergraduate	74 (47.2)	65 (39.9)	75 (54.3)	214 (46.7)	
Tertiary postgraduate	47 (29.9)	57 (35.0)	39 (28.3)	143 (31.3)	
Ethnicity, n (%)					.34
White	124 (79.0)	118 (72.4)	110 (79.7)	352 (76.9)	
Maori/Pacific Islander	12 (7.6)	16 (9.8)	7 (5.1)	35 (7.6)	
Asian	5 (3.2)	11 (6.7)	2 (1.4)	18 (3.9)	
Black African/American	4 (2.5)	3 (1.8)	3 (2.2)	10 (2.2)	
Indigenous	0 (0)	0 (0)	1 (0.7)	1 (0.2)	
Spanish, Hispanic, Latino	5 (3.2)	3 (1.8)	4 (2.9)	12 (2.6)	
Other	7 (4.5)	12 (7.5)	11 (8.0)	30 (6.6)	
Outcome measures, mean (SD)					
Mental Health	64.4 (17.3)	65.7 (17.5)	64.6 (16.5)	64.9 (17.1)	>.99
Vitality	58.5 (17.4)	60.2 (17.4)	58.6 (17.2)	59.1 (17.3)	.59
Depression	3.8 (3.7)	4.0 (4.1)	3.9 (3.5)	3.9 (3.8)	.50
Anxiety	2.8 (2.9)	2.5 (2.6)	2.5 (2.8)	2.6 (2.7)	.83
Stress	5.7 (3.6)	5.7 (3.6)	5.9 (3.4)	5.7 (3.5)	.67
Flourishing	45.2 (7.1)	45.1 (6.8)	44.5 (6.3)	45.0 (6.7)	.10
Satisfaction With Life	23.4 (6.9)	22.5 (7.3)	22.5 (7.0)	22.8 (7.1)	.73

Measurement—Outcomes

Significant within-group improvements were recorded from pre- to postintervention in all groups and across all domains of mental well-being measured (Table 3). Within-group improvements ranged from medium to large effect sizes (Cohen *d*), with flourishing (*d*=0.64), mental health (*d*=0.67), and vitality (*d*=0.74) demonstrating the largest effect. No significant between-group differences were observed for any of the outcome measures (Table 3).

Table 3. Pre- to postchanges in outcome measures of participants and between-group differences.

Outcome Measure	N	Pre, mean (SD)	Post, mean (SD)	Mean change	Change (%)	Within-group change, *P* value	Within-group change, Cohen *d*	Between-group difference, *P* value
Mental health								.77
Gr 1 standard—Emails	103	65.2 (16.4)	74.8 (15.2)	9.6	14.7	<.001	0.72	
Gr 2 Standard plus pSMS[a]	114	67.3 (16.7)	76.1 (13.3)	8.8	13.1	<.001	0.69	
Gr 3 Standard plus VCS[b]	103	66.0 (15.3)	74.2 (14.9)	8.2	12.4	<.001	0.61	
Combined	320	66.2 (16.1)	75.1 (14.4)	8.9	13.4	<.001	0.67	
Vitality								.65
Gr 1 Standard—Emails	103	58.7 (16.8)	69.6 (15.8)	10.9	18.6	<.001	0.78	
Gr 2 Standard plus pSMS	114	61.8 (16.5)	71.9 (13.7)	10.1	16.3	<.001	0.81	
Gr 3 Standard plus VCS	103	60.1 (16.1)	69.2 (15.4)	9.1	15.1	<.001	0.63	
Combined	320	60.3 (16.5)	70.3 (15.0)	10.0	16.6	<.001	0.74	
Depression								.93
Gr 1 Standard—Emails	103	3.5 (3.4)	2.2 (2.6)	−1.3	−37.1	<.001	0.44	
Gr 2 Standard plus pSMS	114	3.5 (3.5)	2.1 (2.7)	−1.4	−40.0	<.001	0.51	
Gr 3 Standard plus VCS	103	3.6 (3.3)	2.2 (2.7)	−1.4	−38.9	<.001	0.50	
Combined	320	3.5 (3.4)	2.2 (2.7)	−1.3	−37.1	<.001	0.48	
Anxiety								.25
Gr 1 Standard—Emails	103	2.7 (2.6)	1.5 (1.8)	−1.2	−44.4	<.001	0.52	
Gr 2 Standard plus pSMS	114	2.1 (2.1)	1.4 (1.8)	−0.7	−33.3	<.001	0.35	
Gr 3 Standard plus VCS	103	2.2 (2.5)	1.3 (1.6)	−0.9	−40.9	<.001	0.43	
Combined	320	2.3 (2.4)	1.4 (1.7)	−0.9	−39.1	<.001	0.43	
Stress								.57
Gr 1 Standard—Emails	103	5.8 (3.5)	4.3 (3.1)	−1.5	−25.9	<.001	0.46	
Gr 2 Standard plus pSMS	114	5.5 (3.4)	4.1 (3.0)	−1.4	−25.5	<.001	0.47	
Gr 3 Standard plus VCS	103	6.0 (3.2)	4.2 (3.0)	−1.8	−30.0	<.001	0.60	
Combined	320	5.7 (3.4)	4.2 (3.0)	−1.5	−26.3	<.001	0.51	
Satisfaction With Life								.65
Gr 1 Standard – Emails	103	23.2 (6.8)	25.6 (6.3)	2.4	10.3	<.001	0.50	
Gr 2 Standard plus pSMS	114	23.1 (7.2)	26.0 (6.6)	2.9	12.6	<.001	0.64	
Gr 3 Standard plus VCS	103	22.9 (6.9)	25.5 (6.5)	2.6	11.4	<.001	0.57	
Combined	320	23.1 (6.9)	25.7 (6.5)	2.6	11.3	<.001	0.58	
Flourishing								.99
Gr 1 Standard—Emails	103	45.2 (6.7)	48.1 (5.0)	2.9	6.4	<.001	0.61	
Gr 2 Standard plus pSMS	114	45.1 (7.0)	48.1 (6.1)	3.0	6.7	<.001	0.65	
Gr 3 Standard plus VCS	103	44.8 (6.2)	47.8 (5.9)	3.0	6.7	<.001	0.64	
Combined	320	45.0 (6.6)	48.0 (5.7)	3.0	6.7	<.001	0.64	

[a]pSMS: personalized SMS.

[b]VCS: videoconferencing support.

Measurement—Adherence

Adherence was not significantly different between groups for mean videos watched (*P*=.42) or mean total challenge points scored (*P*=.71). However, there was notable variability in responses as indicated by the large SDs: videos watched out of 10 (S=6.05 (SD 4.0), S+pSMS=6.48 (SD 3.9), S+VCS=6.75

(SD 3.8); challenge scores out of 1000 (S=369 (SD 362), S+pSMS=340 (SD 339), S+VCS=377 (SD 354).

In the S+VCS group, mean VCSS attendance was 2.8 out of 10, and 37 out of 103 participants (35.9%) had zero attendance. Just 19 out of 103 participants (18.4%) attended 7 or more VCSSs. Secondary analysis revealed participants who attended more than 7 VCSSs, compared with those who attended 6 or less, demonstrated significantly greater improvements in the measures of mental health (P=.006; d=0.71), vitality (P=.005; d=0.73) depression (P=.04; d=0.54), and satisfaction with life (P=.046; d=0.50).

Discussion

Principal Findings

This study compared the influence of 3 modes of human support on the outcomes of a Web- and mobile app–based, lifestyle-focused mental health intervention for a healthy adult cohort. Significant improvements in all domains of mental well-being were recorded in all groups, but the mode of human support had no effect. However, attendance at the VCSSs was low, hindering the ability to draw comparisons.

The study population could be classified as healthy (ie, normal), as evidenced by baseline DASS scores that were within the normal range: depression, mean=3.5 (normal 0-4); anxiety, mean=2.1 (normal 0-3); stress, mean=5.7 (normal 0-7). Notably, despite the "healthy" starting point, medium to large effect size improvements were observed. The results of this study are similar to 2 pilot trials [24,25] and an RCT [26] using the same intervention. Medium to large effects may be due, at least in part, to the multimodal nature of the intervention producing a compounding effect [24]. As the intervention embeds a combination of evidence-based strategies from lifestyle medicine and positive psychology for improving mental well-being, the overall effect of the intervention could be expected to exceed that of a single modality approach.

Videoconferencing has been successfully used as a form of support in various settings, but the sessions were not well attended in this study. Previous research highlights the benefits of videoconferencing as a feasible and acceptable mode of support as long as technology assistance is provided [63]. Benefits include enhanced group bonding, peer observation, personal sharing, and a higher social presence when compared with other forms of digital support [63]. However, although videoconferencing has been used as an effective method to support caregivers [64], chronic disease patients [65], new mothers [66], and postsurgical patients [67], it was underutilized in this intervention.

Several factors may have contributed to the underutilization of the VCSSs. First, Web-based interventions are essentially "pull" technologies, relying on the participant to initiate access to the intervention [110]. Conversely, human support features generally "push" participants to engage with a digital intervention [110]. Within this study, automated emails and SMS messages served as "push" strategies, requiring no effort on the part of the participant. However, videoconferencing support is a "pull" device that required participants to actively seek engagement by attending the scheduled session, and this may have been a contributing factor in low engagement. Second, to overcome potential scheduling barriers, 9 sessions were offered within a week, but this did not translate into higher levels of participation and may have negatively impacted development of group dynamics and peer interaction. Participants could choose to attend any of the 9 sessions offered, which meant group bonding, a known videoconference advantage [63], may have been impeded by a lack of continuity in attendance within each time slot. Notwithstanding, a myriad of other factors may have negatively impacted VCSS engagement, such as time constraints, confidence to use technology, privacy or exposure concerns, and perceptions about effectiveness [63,111]. Undoubtedly, adherence is dynamic in nature [56], and many interindividual variations are still unexplained [49]. The low attendance at the VCSSs highlights the need to investigate engagement facilitators and barriers more thoroughly. Given that the majority of the S+VCS group participants (82%) did not engage regularly with the support provided, this group, in effect, received a comparable level of support with the S group (automated emails), hampering between-group comparability.

Stratified within-group analysis in the S+VCS group showed significantly greater improvement in depression, mental health, vitality, and satisfaction with life metrics for those who attended 7 or more VCSSs. Nevertheless, the results need to be treated with caution because the stratified subgroup is self-selected, no longer randomized, and therefore subject to bias. In addition, the small number of participants and unknown contributing factors (eg, motivation) make drawing conclusions problematic. Elucidating the reasons why participants chose to engage, or not, with videoconferencing support would be an important topic for further research.

Despite the lack of influence of human support in this study, the multimodal intervention demonstrated statistically significant improvement across all groups in all outcome measures, with medium to large effect sizes. Strengthened by results of previous pilot studies [24,25] and the RCT in 2017 [26], it may be feasible to trial the lifestyle-based intervention in primary prevention or clinical settings in the future to attenuate symptoms for those who are at high risk or already suffering from a disorder.

Strengths and Limitations

The intervention, Web design, and additional human support used in this study were underpinned by theoretical models, including the TPB [112], a framework for PSD [93], and key principles of the SAM [34]. Consistency in the VCSSs was enhanced by using 1 experienced, online group facilitator rather than multiple facilitators of varying skill levels. In addition, the study population included a large homogenous cohort, and this was further strengthened by a wide range of age groups from 18 to 81 years.

Limitations included low attendance at VCSSs, which meant that many S+VCS group members probably experienced the intervention similarly to the S group (automated emails), negatively impacting the ability to make meaningful between-group comparisons. Notably, the previous RCT in

2017 [26] demonstrated similar statistically significant improvements within the intervention group to what was seen within the 3 groups of this study. Both the RCT and this study commenced with cohorts who were considered "normal" (indicated by the DASS scores). Conceivably, in adding human support, a ceiling effect was observed, and gauging further improvements above and beyond the benefits of the intervention itself were not realistically measurable.

The cohort were predominantly Seventh-day Adventist church members, which diminishes generalizability to other population groups because of commonly held lifestyle practices (eg, no alcohol or tobacco). Furthermore, participants were skewed toward white females and those who held a tertiary qualification. Although such demographics are commonly portrayed in digital interventions [113,114], these factors limit the generalizability of findings to the broader population. Participants were unblinded, and the study relied on self-reporting for all instruments, which comes with risk of reporting biases (eg, poor recall), and participants are sometimes unaware of their own personal motives and behavior [56].

Other limitations include the failure to measure how participants engaged with automated emails or SMS messages. In addition, we did not gather data regarding the proportion of participants who accessed the program using the mobile app as an alternative to the Web-based platform, which would have been a useful comparison. In future applications, a postintervention survey should include questions regarding the use and preferences for the delivery systems (ie, computer or mobile app), plus participant perceptions regarding the influence of and personal engagement with email and SMS messaging. It would also be important to administer the intervention to a broader population sample to improve generalizability.

Conclusions

The findings of this study strengthen the rationale for Web- and mobile app–based interventions as easily accessible and scalable mental health promotion initiatives. The study demonstrated that a lifestyle-focused intervention improved the mental well-being of a healthy cohort, irrespective of human support. Low VCSS attendance reduced the ability to draw meaningful between-group comparisons, and hence, the influence of different human support modes, although SMS messages provided no significant benefit over automated email support. Stratified analysis demonstrated that regular attendance at a VCSS might be a possible method to enhance outcomes; however, more research is needed regarding factors that influence engagement with that mode of support.

Acknowledgments

The authors would like to thank the volunteer participants, the Web development team, and the Seventh-day Adventist Church Health department of the South Pacific Division that provided technical support and free access to the intervention on the eLMS and mobile app.

Conflicts of Interest

GP is employed by the South Pacific Division of the Seventh-day Adventist Church, which promotes and directs the intervention among members and the wider community. DM administers a "profit-for-purpose" Trust using a version of the intervention; he receives no personal financial remuneration. No authors have a financial interest in the intervention and no other authors have conflicts of interest to declare.

Multimedia Appendix 1
Website and app screenshots.
[PPTX File , 4354 KB - jmir_v22i1e15592_app1.pptx]

Multimedia Appendix 2
Timetable overview.
[DOCX File , 14 KB - jmir_v22i1e15592_app2.docx]

Multimedia Appendix 3
Videoconference meeting schedule.
[PDF File (Adobe PDF File), 178 KB - jmir_v22i1e15592_app3.pdf]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 283 KB - jmir_v22i1e15592_app4.pdf]

References

1. World Health Organization. 2013. Comprehensive Mental Health Action Plan 2013-2020 URL: https://www.who.int/mental_health/action_plan_2013/en/ [accessed 2019-07-17]

2. World Health Organization. 2018. Mental Disorders URL: https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders [accessed 2019-07-17]
3. World Health Organization. 2018. Depression URL: https://www.who.int/en/news-room/fact-sheets/detail/depression [accessed 2019-07-11]
4. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017 Feb;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
5. Stephenson CP, Karanges E, McGregor IS. Trends in the utilisation of psychotropic medications in Australia from 2000 to 2011. Aust N Z J Psychiatry 2013 Jan;47(1):74-87. [doi: 10.1177/0004867412466595] [Medline: 23144164]
6. O'Neil A, Jacka FN, Quirk SE, Cocker F, Taylor CB, Oldenburg B, et al. A shared framework for the common mental disorders and Non-Communicable Disease: key considerations for disease prevention and control. BMC Psychiatry 2015 Feb 5;15:15 [FREE Full text] [doi: 10.1186/s12888-015-0394-0] [Medline: 25652365]
7. Finley CR, Chan DS, Garrison S, Korownyk C, Kolber MR, Campbell S, et al. What are the most common conditions in primary care? Systematic review. Can Fam Physician 2018 Nov;64(11):832-840 [FREE Full text] [Medline: 30429181]
8. Australian Psychological Society. Heads Up. Stress & Wellbeing: How Australians Are Coping With Life URL: https://www.headsup.org.au/docs/default-source/default-document-library/stress-and-wellbeing-in-australia-report.pdf?sfvrsn=7f08274d_4 [accessed 2017-07-25]
9. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2015 Apr;22(e1):e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
10. 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]
11. Hintz S, Frazier PA, Meredith L. Evaluating an online stress management intervention for college students. J Couns Psychol 2015 Apr;62(2):137-147. [doi: 10.1037/cou0000014] [Medline: 24635586]
12. Bhayee S, Tomaszewski P, Lee DH, Moffat G, Pino L, Moreno S, et al. Attentional and affective consequences of technology supported mindfulness training: a randomised, active control, efficacy trial. BMC Psychol 2016 Nov 29;4(1):60 [FREE Full text] [doi: 10.1186/s40359-016-0168-6] [Medline: 27894358]
13. Keng S, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
14. Galante J, Dufour G, Vainre M, Wagner AP, Stochl J, Benton A, et al. A mindfulness-based intervention to increase resilience to stress in university students (the Mindful Student Study): a pragmatic randomised controlled trial. Lancet Public Health 2018 Feb;3(2):e72-e81 [FREE Full text] [doi: 10.1016/S2468-2667(17)30231-1] [Medline: 29422189]
15. Wahbeh H, Oken BS. Internet mindfulness meditation intervention for the general public: pilot randomized controlled trial. JMIR Ment Health 2016 Aug 8;3(3):e37 [FREE Full text] [doi: 10.2196/mental.5900] [Medline: 27502759]
16. Blanchflower DG, Oswald AJ, Stewart-Brown S. Is psychological well-being linked to the consumption of fruit and vegetables? Soc Indic Res 2012 Oct;114(3):785-801. [doi: 10.1007/s11205-012-0173-y]
17. White BA, Horwath CC, Conner TS. Many apples a day keep the blues away--daily experiences of negative and positive affect and food consumption in young adults. Br J Health Psychol 2013 Nov;18(4):782-798. [doi: 10.1111/bjhp.12021] [Medline: 23347122]
18. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med 2013 Nov;45(5):649-657. [doi: 10.1016/j.amepre.2013.08.001] [Medline: 24139780]
19. Stubbs B, Vancampfort D, Rosenbaum S, Ward PB, Richards J, Ussher M, et al. Challenges establishing the efficacy of exercise as an antidepressant treatment: a systematic review and meta-analysis of control group responses in exercise randomised controlled trials. Sports Med 2016 May;46(5):699-713. [doi: 10.1007/s40279-015-0441-5] [Medline: 26707338]
20. Lyubomirsky S, Dickerhoof R, Boehm JK, Sheldon KM. Becoming happier takes both a will and a proper way: an experimental longitudinal intervention to boost well-being. Emotion 2011 Apr;11(2):391-402 [FREE Full text] [doi: 10.1037/a0022575] [Medline: 21500907]
21. Proyer RT, Gander F, Wellenzohn S, Ruch W. Positive psychology interventions in people aged 50-79 years: long-term effects of placebo-controlled online interventions on well-being and depression. Aging Ment Health 2014;18(8):997-1005. [doi: 10.1080/13607863.2014.899978] [Medline: 24712501]
22. Gander F, Proyer RT, Ruch W. Positive psychology interventions addressing pleasure, engagement, meaning, positive relationships, and accomplishment increase well-being and ameliorate depressive symptoms: a randomized, placebo-controlled online study. Front Psychol 2016;7:686 [FREE Full text] [doi: 10.3389/fpsyg.2016.00686] [Medline: 27242600]
23. Morton DP. Combining Lifestyle Medicine and Positive Psychology to Improve Mental Health and Emotional Well-being. Am J Lifestyle Med 2018;12(5):370-374 [FREE Full text] [doi: 10.1177/1559827618766482] [Medline: 30283261]
24. Morton DP, Hinze J, Craig B, Herman W, Kent L, Beamish P, et al. A multimodal intervention for improving the mental health and emotional well-being of college students. Am J Lifestyle Med 2017. [doi: 10.1177/1559827617733941]
25. Hinze J, Morton D. Wellbeing education for educators. TEACH J Christian Educ 2017;11(2):50-57 [FREE Full text]

26. Przybylko G, Morton D, Kent L, Morton J, Hinze J, Beamish P, et al. The effect of an online multimodal lifestyle intervention on mental health and emotional wellness: A randomised control trial. 2018 Presented at: Poster presented at the American College of Lifestyle Medicine Conference; Oct 21-24; Indianapolis, Indiana.
27. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, Campayo JG, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
28. Woodworth RJ, O'Brien-Malone A, Diamond MR, Schüz B. Web-based positive psychology interventions: a reexamination of effectiveness. J Clin Psychol 2017 Mar;73(3):218-232. [doi: 10.1002/jclp.22328] [Medline: 27377826]
29. Choi I, Sharpe L, Li S, Hunt C. Acceptability of psychological treatment to Chinese- and Caucasian-Australians: internet treatment reduces barriers but face-to-face care is preferred. Soc Psychiatry Psychiatr Epidemiol 2015 Jan;50(1):77-87. [doi: 10.1007/s00127-014-0921-1] [Medline: 24993290]
30. 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;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
31. 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]
32. Ebert DD, van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
33. Shim M, Mahaffey B, Bleidistel M, Gonzalez A. A scoping review of human-support factors in the context of Internet-based psychological interventions (IPIs) for depression and anxiety disorders. Clin Psychol Rev 2017 Nov;57:129-140. [doi: 10.1016/j.cpr.2017.09.003] [Medline: 28934623]
34. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
35. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interv 2014;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
36. 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]
37. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
38. 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]
39. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
40. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 1;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
41. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
42. 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]
43. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided . self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
44. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for Social Anxiety Disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
45. Garland EL, Fredrickson B, Kring AM, Johnson DP, Meyer PS, Penn DL. Upward spirals of positive emotions counter downward spirals of negativity: insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. Clin Psychol Rev 2010 Nov;30(7):849-864 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.002] [Medline: 20363063]

46. Christensen H, Batterham P, Mackinnon A, Griffiths KM, Hehir KK, Kenardy J, et al. Prevention of generalized anxiety disorder using a web intervention, iChill: randomized controlled trial. J Med Internet Res 2014 Sep 2;16(9):e199 [FREE Full text] [doi: 10.2196/jmir.3507] [Medline: 25270886]
47. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized clinical trial of a primary care internet-based intervention to prevent adolescent depression: one-year outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
48. Aardoom JJ, Dingemans AE, Spinhoven P, van Ginkel JR, de Rooij M, van Furth EF. Web-based fully automated self-help with different levels of therapist support for individuals with eating disorder symptoms: a randomized controlled trial. J Med Internet Res 2016 Jun 17;18(6):e159 [FREE Full text] [doi: 10.2196/jmir.5709] [Medline: 27317358]
49. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
50. Allexandre D, Bernstein AM, Walker E, Hunter J, Roizen MF, Morledge TJ. A web-based mindfulness stress management program in a corporate call center: a randomized clinical trial to evaluate the added benefit of onsite group support. J Occup Environ Med 2016 Mar;58(3):254-264 [FREE Full text] [doi: 10.1097/JOM.0000000000000680] [Medline: 26949875]
51. 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]
52. Lancee J, van den Bout J, Sorbi MJ, van Straten A. Motivational support provided via email improves the effectiveness of internet-delivered self-help treatment for insomnia: a randomized trial. Behav Res Ther 2013 Dec;51(12):797-805. [doi: 10.1016/j.brat.2013.09.004] [Medline: 24121097]
53. Stefanopoulou E, Lewis D, Taylor M, Broscombe J, Ahmad J, Larkin J. Are digitally delivered psychological interventions for depression the way forward? A review. Psychiatr Q 2018 Dec;89(4):779-794. [doi: 10.1007/s11126-018-9576-5] [Medline: 29656348]
54. Kelders SM, Bohlmeijer ET, Pots WT, van Gemert-Pijnen JE. Comparing human and automated support for depression: fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]
55. Titov N, Andrews G, Choi I, Schwencke G, Johnston L. Randomized controlled trial of web-based treatment of social phobia without clinician guidance. Aust N Z J Psychiatry 2009;43(10):913-919 [FREE Full text] [doi: 10.1080/00048670903179160]
56. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
57. 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]
58. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet J, Alessi M, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]
59. Strandbygaard U, Thomsen SF, Backer V. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. Respir Med 2010 Feb;104(2):166-171 [FREE Full text] [doi: 10.1016/j.rmed.2009.10.003] [Medline: 19854632]
60. Akhu-Zaheya LM, Shiyab WY. The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases. Int J Med Inform 2017 Feb;98:65-75. [doi: 10.1016/j.ijmedinf.2016.12.003] [Medline: 28034414]
61. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
62. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable Sensor/Device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
63. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 2;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
64. Damianakis T, Tough A, Marziali E, Dawson DR. Therapy online: a web-based video support group for family caregivers of survivors with traumatic brain injury. J Head Trauma Rehabil 2016;31(4):E12-E20. [doi: 10.1097/HTR.0000000000000178] [Medline: 26291634]
65. Marziali E. E-health program for patients with chronic disease. Telemed J E Health 2009 Mar;15(2):176-181. [doi: 10.1089/tmj.2008.0082] [Medline: 19292627]

66. Nyström K, Ohrling K. Parental support: mothers' experience of electronic encounters. J Telemed Telecare 2006;12(4):194-197. [doi: 10.1258/135763306777488726] [Medline: 16774701]
67. Wild B, Hünnemeyer K, Sauer H, Hain B, Mack I, Schellberg D, et al. A 1-year videoconferencing-based psychoeducational group intervention following bariatric surgery: results of a randomized controlled study. Surg Obes Relat Dis 2015;11(6):1349-1360. [doi: 10.1016/j.soard.2015.05.018] [Medline: 26421929]
68. Carrère S, Buehlman KT, Gottman JM, Coan JA, Ruckstuhl L. Predicting marital stability and divorce in newlywed couples. J Fam Psychol 2000 Mar;14(1):42-58. [Medline: 10740681]
69. Losada M. The complex dynamics of high performance teams. Math Comput Model 1999;30(9-10):179-192 [FREE Full text] [doi: 10.1016/s0895-7177(99)00189-2]
70. Nair S, Sagar M, Sollers J, Consedine N, Broadbent E. Do slumped and upright postures affect stress responses? A randomized trial. Health Psychol 2015 Jun;34(6):632-641. [doi: 10.1037/hea0000146] [Medline: 25222091]
71. Josefsson T, Lindwall M, Archer T. Physical exercise intervention in depressive disorders: meta-analysis and systematic review. Scand J Med Sci Sports 2014 Apr;24(2):259-272. [doi: 10.1111/sms.12050] [Medline: 23362828]
72. Lathia N, Sandstrom GM, Mascolo C, Rentfrow PJ. Happier people live more active lives: using smartphones to link happiness and physical activity. PLoS One 2017;12(1):e0160589 [FREE Full text] [doi: 10.1371/journal.pone.0160589] [Medline: 28052069]
73. Richards J, Jiang X, Kelly P, Chau J, Bauman A, Ding D. Don't worry, be happy: cross-sectional associations between physical activity and happiness in 15 European countries. BMC Public Health 2015 Jan 31;15:53 [FREE Full text] [doi: 10.1186/s12889-015-1391-4] [Medline: 25636787]
74. Barton J, Pretty J. What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. Environ Sci Technol 2010 May 15;44(10):3947-3955. [doi: 10.1021/es903183r] [Medline: 20337470]
75. Mantler A, Logan AC. Natural environments and mental health. Adv Integr Med 2015;2(1):5-12 [FREE Full text] [doi: 10.1016/j.aimed.2015.03.002]
76. McMahan E, Estes D. The effect of contact with natural environments on positive and negative affect: a meta-analysis. J Posit Psychol 2015;10(6):507-519 [FREE Full text] [doi: 10.1080/17439760.2014.994224]
77. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. Br Med J 2008 Dec 4;337:a2338 [FREE Full text] [doi: 10.1136/bmj.a2338] [Medline: 19056788]
78. Toussaint LL, Worthington E, Williams DR, editors. Forgiveness and Health Scientific Evidence and Theories Relating Forgiveness to Better Health. New York: Springer; 2015.
79. Toussaint L, Shields GS, Dorn G, Slavich GM. Effects of lifetime stress exposure on mental and physical health in young adulthood: How stress degrades and forgiveness protects health. J Health Psychol 2016 Jun;21(6):1004-1014 [FREE Full text] [doi: 10.1177/1359105314544132] [Medline: 25139892]
80. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003 Feb;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377] [Medline: 12585811]
81. 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]
82. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
83. Jacka FN, O'Neil A, Opie R, Itsiopoulos C, Cotton S, Mohebbi M, et al. A randomised controlled trial of dietary improvement for adults with major depression (the 'SMILES' trial). BMC Med 2017 Jan 30;15(1):23 [FREE Full text] [doi: 10.1186/s12916-017-0791-y] [Medline: 28137247]
84. Quirk SE, Williams LJ, O'Neil A, Pasco JA, Jacka FN, Housden S, et al. The association between diet quality, dietary patterns and depression in adults: a systematic review. BMC Psychiatry 2013 Jun 27;13:175 [FREE Full text] [doi: 10.1186/1471-244X-13-175] [Medline: 23802679]
85. Neckelmann D, Mykletun A, Dahl AA. Chronic insomnia as a risk factor for developing anxiety and depression. Sleep 2007 Jul;30(7):873-880 [FREE Full text] [doi: 10.1093/sleep/30.7.873] [Medline: 17682658]
86. Bedrosian TA, Nelson RJ. Influence of the modern light environment on mood. Mol Psychiatry 2013 Jul;18(7):751-757. [doi: 10.1038/mp.2013.70] [Medline: 23711982]
87. Cornélissen G, Watson D, Mitsutake G, Fišer B, Siegelová J, Dušek J, Vohlídalová, et al. Mapping of circaseptan and circadian changes in mood. Scr Med (Brno) 2005;78(2):89-98 [FREE Full text] [Medline: 18985163]
88. Smith-Gabai H, Ludwig F. Observing the Jewish sabbath: a meaningful restorative ritual for modern times. J Occup Sci 2011;18(4):347-355 [FREE Full text] [doi: 10.1080/14427591.2011.595891]
89. Buchanan KE, Bardi A. Acts of kindness and acts of novelty affect life satisfaction. J Soc Psychol 2010;150(3):235-237. [doi: 10.1080/00224540903365554] [Medline: 20575332]
90. Seligman ME. Flourish: A Visionary New Understanding of Happiness and Well-being. New York: Free Press; 2013.
91. Tkach CT. Unlocking the treasury of human kindness: enduring improvements in mood, happiness, and self-evaluations. Diss Abstr Int 2006;67(1-B):603 [FREE Full text]

92. Morton D. Live More Happy: Scientifically Proven Ways to Lift Your Mood and Your Life. Warburton, Australia: Signs Publishing; 2018.
93. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inform Syst 2009;24(1):485-500. [doi: 10.17705/1CAIS.02428]
94. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco: Elsevier Science; 2003.
95. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
96. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Br Med J 1992 Jul 18;305(6846):160-164 [FREE Full text] [doi: 10.1136/bmj.305.6846.160] [Medline: 1285753]
97. Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. Qual Life Res 1994 Feb;3(1):7-12. [doi: 10.1007/bf00647843] [Medline: 8142947]
98. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66. [doi: 10.1097/00005650-199401000-00004] [Medline: 8277801]
99. 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]
100. Randall D, Thomas M, Whiting D, McGrath A. Depression Anxiety Stress Scales (DASS-21): factor structure in traumatic brain injury rehabilitation. J Head Trauma Rehabil 2017;32(2):134-144. [doi: 10.1097/HTR.0000000000000250] [Medline: 27455437]
101. Le MT, Tran TD, Holton S, Nguyen HT, Wolfe R, Fisher J. Reliability, convergent validity and factor structure of the DASS-21 in a sample of Vietnamese adolescents. PLoS One 2017;12(7):e0180557 [FREE Full text] [doi: 10.1371/journal.pone.0180557] [Medline: 28723909]
102. Osman A, Wong JL, Bagge CL, Freedenthal S, Gutierrez PM, Lozano G. The Depression Anxiety Stress Scales-21 (DASS-21): further examination of dimensions, scale reliability, and correlates. J Clin Psychol 2012 Dec;68(12):1322-1338. [doi: 10.1002/jclp.21908] [Medline: 22930477]
103. Diener E, Emmons R, Larsen R, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
104. Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction with Life Scale: evidence for the cross-method convergence of well-being measures. J Pers Assess 1991 Aug;57(1):149-161. [doi: 10.1207/s15327752jpa5701_17] [Medline: 1920028]
105. Pavot WD, Diener E. Review of the satisfaction with life scale. Psychol Assess 1993;5(2):164-172 [FREE Full text] [doi: 10.1037//1040-3590.5.2.164]
106. Vassar M. A note on the score reliability for the Satisfaction With Life Scale: an RG study. Soc Indic Res 2008;86(1):47-57 [FREE Full text] [doi: 10.1007/s11205-007-9113-7]
107. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2010;97(2):143-156 [FREE Full text] [doi: 10.1007/s11205-009-9493-y]
108. Silva AJ, Caetano A. Validation of the flourishing scale and scale of positive and negative experience in Portugal. Soc Indic Res 2013;110(2):469-478 [FREE Full text] [doi: 10.1007/s11205-011-9938-y]
109. 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]
110. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
111. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
112. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control. Berlin, Heidelberg: Springer; 1985:11-39.
113. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Comput Hum Behav 2009;25(3):749-760 [FREE Full text] [doi: 10.1016/j.chb.2009.02.003]
114. Waters LA, Galichet B, Owen N, Eakin E. Who participates in physical activity intervention trials? J Phys Act Health 2011 Jan;8(1):85-103. [Medline: 21297189]

Abbreviations

7DWI: 7 Dimensions of Well-Being Index

CONSORT: Consolidated Standards of Reporting Trials
DASS: Depression, Anxiety and Stress Scale
eLMS: electronic learning management system
MHPI: mental health promotion intervention
PSD: persuasive systems design
RCT: randomized controlled trial
S: standard group
S+pSMS: standard plus personalized short message service group
S+VCS: standard plus videoconferencing support group
SAM: supportive accountability model
SF-36: Short Form Health Survey
SWLS: Satisfaction With Life Scale
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 24.07.19; peer-reviewed by JR Bautista, C Salewski, S Poduval; comments to author 21.09.19; revised version received 14.10.19; accepted 20.10.19; published 06.01.20.

Please cite as:
Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, Craig BA
A Web- and Mobile App–Based Mental Health Promotion Intervention Comparing Email, Short Message Service, and Videoconferencing Support for a Healthy Cohort: Randomized Comparative Study
J Med Internet Res 2020;22(1):e15592
URL: https://www.jmir.org/2020/1/e15592
doi:10.2196/15592
PMID:31904578

Original Paper

Efficacy of a Self-Help Web-Based Recovery Training in Improving Sleep in Workers: Randomized Controlled Trial in the General Working Population

Doerte Behrendt[1], MSc; David Daniel Ebert[2], PhD; Kai Spiegelhalder[3], PhD; Dirk Lehr[1], PhD

[1]Department of Health Psychology and Applied Biological Psychology, Institute of Psychology, Leuphana University of Lueneburg, Lueneburg, Germany

[2]Faculty of Behavioural and Movement Sciences, Clinical, Neuro- & Developmental Psychology, Free University Amsterdam, Amsterdam, Germany

[3]Department of Psychiatry and Psychotherapy, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Corresponding Author:
Doerte Behrendt, MSc
Department of Health Psychology and Applied Biological Psychology
Institute of Psychology
Leuphana University of Lueneburg
Universitätsallee 1
Lueneburg, 21335
Germany
Phone: 49 41316772374
Email: behrendt@leuphana.de

Abstract

Background: Sleep complaints are among the most prevalent health concerns, especially among workers, which may lead to adverse effects on health and work. Internet-delivered cognitive behavioral therapy for insomnia (iCBT-I) offers the opportunity to deliver effective solutions on a large scale. The efficacy of iCBT-I for clinical samples has been demonstrated in recent meta-analyses, and there is evidence that iCBT-I is effective in the working population with severe sleep complaints. However, to date, there is limited evidence from randomized controlled trials that iCBT-I could also be an effective tool for universal prevention among the general working population regardless of symptom severity. Although increasing evidence suggests that negatively toned cognitive activity may be a key factor for the development and maintenance of insomnia, little is known about how iCBT-I improves sleep by reducing presleep cognitive activity.

Objective: This study aimed to examine the efficacy of a self-help internet-delivered recovery training, based on principles of iCBT-I tailored to the work-life domain, among the general working population. General and work-related cognitive activities were investigated as potential mediators of the intervention's effect.

Methods: A sample of 177 workers were randomized to receive either the iCBT-I (n=88) or controls (n=89). The intervention is a Web-based training consisting of six 1-week modules. As the training was self-help, participants received nothing but technical support via email. Web-based self-report assessments were scheduled at baseline, at 8 weeks, and at 6 months following randomization. The primary outcome was insomnia severity. Secondary outcomes included measures of mental health and work-related health and cognitive activity. In an exploratory analysis, general and work-related cognitive activities, measured as worry and work-related rumination, were investigated as mediators.

Results: Analysis of the linear mixed effects model showed that, relative to controls, participants who received iCBT-I reported significantly lower insomnia severity scores at postintervention (between-group mean difference −4.36; 95% CI −5.59 to − 3.03; Cohen *d*=0.97) and at 6-month follow-up (between-group difference: −3.64; 95% CI −4.89 to −2.39; Cohen *d*=0.86). The overall test of group-by-time interaction was significant (*P*<.001). Significant differences, with small-to-large effect sizes, were also detected for cognitive activity and for mental and work-related health, but not for absenteeism. Mediation analysis demonstrated that work-related rumination (indirect effect: a_1b_1=−0.80; SE=0.34; 95% boot CI −1.59 to −0.25) and worry (indirect effect: a_2b_2=−0.37; SE=0.19; 95% boot CI −0.85 to −0.09) mediate the intervention's effect on sleep.

Conclusions: A self-help Web-based recovery training, grounded in the principles of iCBT-I, can be effective in the general working population, both short and long term. Work-related rumination may be a particularly crucial mediator of the intervention's

effect, suggesting that tailoring interventions to the workplace, including components to reduce the work-related cognitive activity, might be important when designing recovery interventions for workers.

Trial Registration: German Clinical Trials Register DRKS00007142; https://www.drks.de/DRKS00007142

(J Med Internet Res 2020;22(1):e13346) doi:10.2196/13346

KEYWORDS

occupational health; e-mental-health; insomnia; Web-based, cognitive behavioral therapy; mediators

Introduction

Background

Impaired sleep is a common complaint and among the most prevalent health concerns. In Western industrialized countries, approximately 30% to 35% of the general adult population report insomnia symptoms, and approximately 10% meet the criteria of insomnia as a disorder [1]. A similar situation exists in the German working population, among whom 1 study documented a 4-week prevalence of 9.4% for insomnia disorder [2]. In the same study, an additional 35.1% of the working population reported subclinical symptoms of insomnia [2]. The burden of impaired sleep is 2-fold. First, insomnia has a high impact on quality of life and daytime functioning, which includes increasing absenteeism [3]. These effects place workers at higher risk for workplace injuries and may adversely affect their work performance [4]. Second, insomnia is a risk factor for impaired health, including cardiovascular disease [4,5], metabolic syndrome [6], and a variety of mental disorders [7], especially depression [8]. In addition, subclinical insomnia that remains untreated or for which treatment is delayed places individuals at risk for developing clinically significant insomnia. This is important because untreated insomnia is associated with negative long-term health outcomes [9-11].

Published guidelines recommend insomnia-specific cognitive behavioral therapy (CBT-I) as first-line treatment [12], as it has been shown to produce large, sustainable effects [13]. However, CBT-I is not widely available, primarily offered in specialized research settings. Thus, disseminating CBT-I is a major public health challenge. Internet-delivered interventions have been suggested as a potential solution, as they are accessible to a greater number and a broader range of people [14]. Recent meta-analyses involving clinical samples have demonstrated that internet-delivered CBT-I (iCBT-I) is an effective alternative to face-to-face CBT-I [15-18]. Moreover, iCBT-I has been shown to improve not only insomnia but also other mental health outcomes, including symptoms of depression [19]. In addition, iCBT-I appears to be cost-effective from an employer's perspective [20]. Especially when delivered in a self-help format without guidance, internet-based interventions have the potential to provide easy and affordable access to evidence-based interventions to a large population [21].

This said, most previous studies on iCBT-I targeted clinical samples or were conducted in an indicated prevention setting, not having the focus on the working population in general [15-18]. Consequently, results from previously published studies on iCBT-I might not be generalizable to the general working population, in which workers with severe sleep complaints and workers with lower or no sleep complaints are included.

Prior Interventional Studies in the Working Population

Previous studies in the working population give first indications for the efficacy of face-to-face CBT-I for workers [22,23]. Moreover, recently conducted studies evaluating the efficacy of iCBT-I provided evidence that iCBT-I is also effective in improving sleep in the working population in the context of indicated prevention with moderate-to-large effects in the short term [24,25] and prevention with large effects in the long term [26] For specific occupational groups, namely teachers, large effects were found, both short and long-term [20,27,28]. Limitations of the prior studies are that they only included workers with elevated insomnia symptoms, thereby excluding a substantial portion of the working population with less severe insomnia symptoms. To the best of our knowledge, to date, only 2 other randomized controlled trials have evaluated the efficacy of self-help iCBT-I in a general working population. As both the trials include all interested workers regardless of insomnia symptoms, with moderate-to-large effects on sleep in the short term [29,30], they mimic a universal prevention approach.

Hence, there is insufficient evidence on whether iCBT-I might also be effective long term for the general working population without inclusion criteria on insomnia severity. This knowledge is important to clarify the question of whether iCBT-I should be part of universal prevention.

Besides assessing the efficacy of iCBT-I in the general working population, it is also of importance to understand the mechanisms underlying iCBT-I for workers. Considering risk factors for impaired sleep in workers, there is evidence that certain psychosocial work characteristics (eg, high job demands, low job control, and low perceived support) adversely impact sleep [31]. Increasing evidence indicates that negatively toned presleep cognitive activity may be an important mechanism for the relationship between work stressors and sleeping problems [32-36]. One path through which work stressors might impact sleep could be via work-related cognitive activity before sleep (eg, when individuals ruminate about previous problems at work). For example, Berset et al [33] found that work-related rumination is a mediator of work stressors' effects on self-reported sleep quality. This can be explained by cognitive models of insomnia, which highlight the importance of increased negative cognitive activity in the development and maintenance of insomnia [32,37].

Another form of presleep cognitive activity is general worry about events that might occur in the future that are characterized by potential negative outcomes (eg, worrying about the anticipated consequences of sleep loss, similar to impaired

performance the next day) [38]. Previous studies have shown that general worry before sleep plays a mediating role linking stress and impaired sleep [39,40].

Although general cognitive activity and work-related cognitive activity may involve common cognitive processes [41], the content of the 2 forms of cognitive activity is different. Consequently, it might be useful to apply different strategies, thereby targeting each process, when treating insomnia [42].

One finding from a previous study of this present training was that general cognitive activity pertaining to worry mediates the intervention's effect on sleep [28]. However, evidence is missing on whether sleep improvement might also be mediated by reducing work-related cognitive activity [43,44]. As such, it also remains unclear whether adding training elements to reduce work-related cognitive activity and tailoring the intervention to the work-life domain is important and might incrementally contribute to designing interventions to improve sleep among workers.

Aim of the Study

The primary aim of this study was to investigate the efficacy of a self-help version of iCBT-I, which is accessible to all workers who are interested in participating in a training to improve their sleep, employing both short- and long-term assessments of intervention effects.

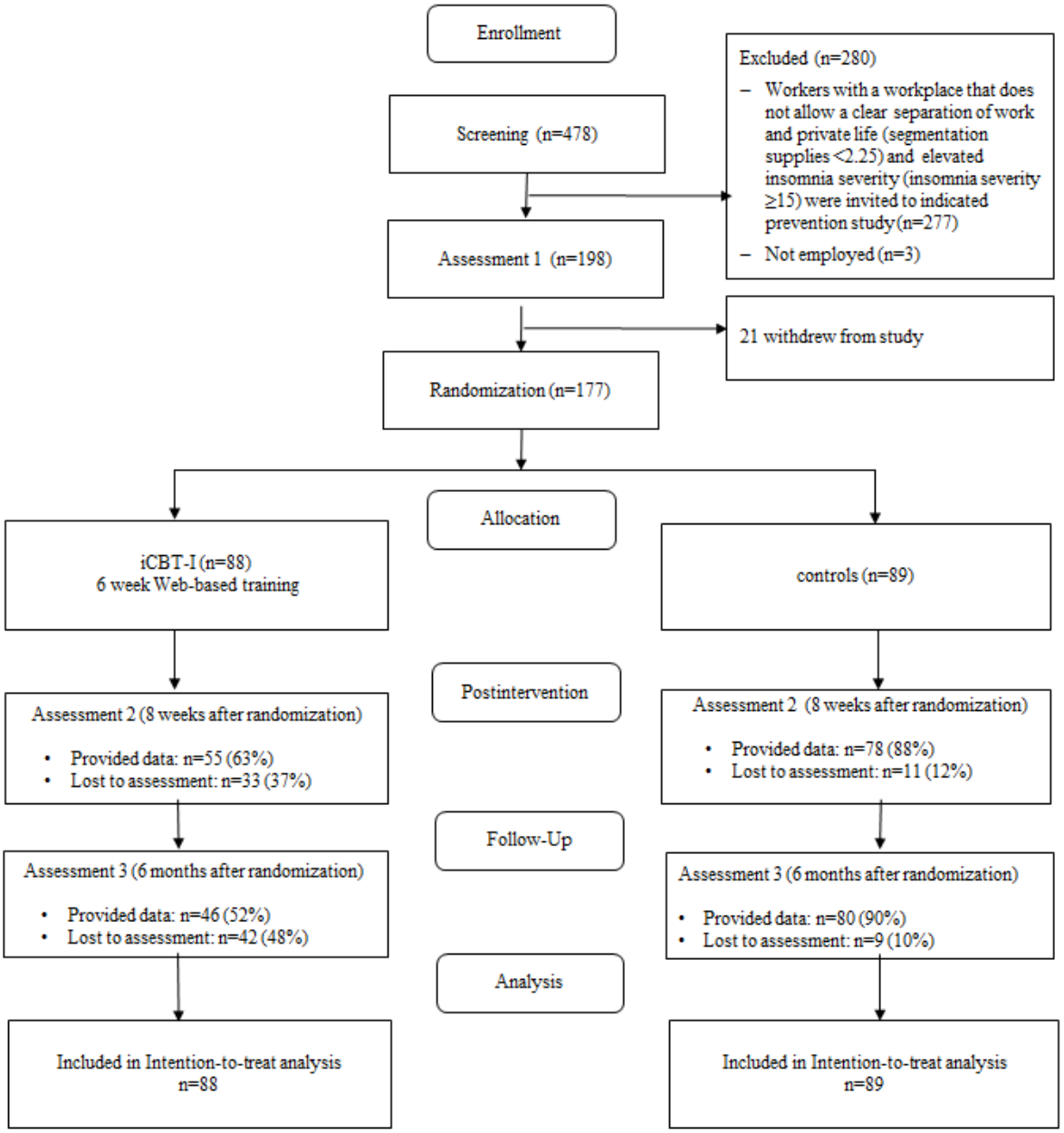
As a secondary aim, we examined whether, in addition to the role of general cognitive activity, work-related cognitive activity is a putative mediator of the intervention's efficacy in workers.

Methods

Study Design and Time Frame

In this randomized trial, subjects were randomly assigned to either receive access to an iCBT-I or received the iCBT-I 6 months later and had in the meantime full access to routine occupational health care. Primary and secondary outcomes were self-assessed Web-based at baseline (T1), postintervention (8 weeks postrandomization, T2), and at 6-month follow-up (4 months postintervention, T3; see Figure 1 for details). The Ethics Committee of Leuphana University of Lueneburg (Lehr201411_Schlaftraining) approved this study that was registered as DRKS00007142 in the German Clinical Trials Register (DRKS).

Figure 1. Flow of study participants. iCBT-I: internet-delivered cognitive behavioral therapy for insomnia.



Sample Size Calculation

Previous studies on this training have yielded intervention effects for insomnia severity ranging from d=1.37 to d=1.45 in teachers with elevated insomnia symptoms. However, studies following a universal health prevention approach without symptom severity as the inclusion criterion are usually associated with smaller effect sizes [45,46]. Thus, our sample size calculation to test the efficacy of the iCBT-I assumed an effect size of d=0.4 for the primary outcome (insomnia severity) at postintervention because of our inclusion of all interested workers regardless of symptom severity. We, therefore, estimated needing 200 participants to achieve 80% power (beta=.80) and 95% CI (alpha=.05) in a 2-tailed analysis.

Inclusion and Exclusion Criteria

All workers who were aged at least 18 years, who were interested in improving their sleep through an iCBT-I, and who had access to the internet were invited to participate in the study. However, workers who expressed suicidal ideations—indicated by their response to item 9 (>1) from the Beck Depression Inventory-II [47]—were excluded. Individuals taking sleep medication were not excluded from the study but were asked to maintain a constant dose and not change their sleep medication for the full course of the study.

Procedures

Recruitment took place via both the media (national telecast and articles) and an email distribution list from the occupational health program provided by one of the largest health insurance companies in Germany. Workers who (1) are motivated to promote their sleep, (2) want to mentally detach from problems

at work, and (3) want to actively recover were addressed. Participants for 2 independent studies were recruited at the same time via the same recruiting channels. The same screening process for these 2 independent trials with different evaluation focus was used. In this trial, we evaluated the efficacy of the training in a general working population sample, where all interested workers could participate regardless of insomnia severity or any other sleep or workplace-specific characteristics.

For the other trial, workers with a workplace that is characterized by blurred boundaries between work and nonwork life (segmentation supplies <2.25) [48] and that showed a score of ≥15 of insomnia severity [49] were recruited (DRKS: DRKS00006223).

Interested workers registered for the study at the landing page of the training and had to provide an email address and a first and last name that could be pseudonyms if desired. Once registered, an individual profile on the Web-based training platform was created. Registered workers received an email with detailed information about the study and a link to the online screening questionnaire on the training platform. All individuals who (1) were employed, (2) had access to the internet, and (2) expressed no suicidal ideations were asked to complete the online baseline assessment and sign the informed consent form, after which they were randomized to 1 of the 2 intervention arms, using a computer-generated randomization list with a ratio of 1:1 and a block size of 2 [50]. The randomization list was generated, and randomization was performed blinded by 2 researchers in our department who were not otherwise involved in this study. Blinding to group allocation was not feasible. The participants were informed about the randomization outcome via email. Participants who were allocated to the training group had immediate access to the intervention. Participants in the control group received access to the training after their 6-month follow-up assessment; in the meantime, they had full access to the usual care offered by routine health care services throughout the trial.

Intervention

In 2 previous studies, a version of the present training that was tailored to teachers was evaluated in school teachers [27,28]. The guided version of the training has been demonstrated to be effective for teachers in both the short term and long term [27]. Meanwhile, evidence supporting efficacy of the self-help version for teachers is restricted to its short-term effects only [28]. For this study, the teacher-specific version of the intervention was revised and adapted for the general working population. The intervention was developed for workers who have recurring problems initiating and maintaining sleep, who tend to ruminate about their work, and who have problems detaching from work. This present training is an internet-delivered training based on CBT-I that was specifically tailored to the needs of those workers, including elements to reduce work-related cognitive activity in the evening, in addition to well-established CBT-I methods, such as techniques to reduce general worry about anticipated consequences of sleep loss [51]. This training can be implemented in either a guided or unguided self-help format.

The training consists of six 1-week modules, each lasting 45 to 60 min. Participants could process the training the way they wanted and could access modules at any time. They were told an ideal completion time of 1 module per week. Self-help iCBT-I is beneficial, in particular, to participants with different levels of sleep complaints, as users can focus on those exercises that seem most helpful for them and can shorten others.

The modules focus on the following subjects: In module 1, subjects are provided psychoeducation on healthy sleep, plan which sleep hygiene rules they are going to follow over the subsequent week, and are introduced to the concept of a sleep diary. Module 2 focuses on sleep restriction and stimulus control. Participants are asked to plan the first step in sleep restriction and reschedule their sleep for the following week accordingly. To support this, each participant could use a Web-based sleep diary that was available on the training platform.

In module 3, participants review their progress on sleep restriction and sleep hygiene and then schedule their sleep for the next week according to their sleep restriction plan. Besides, boundary tactics to support mental detachment from work are targeted; and a gratitude diary is introduced to prevent ruminative thoughts before sleeping. During module 4, participants receive information on work-related rumination and worry, including their impact on sleep, and techniques (eg, relaxation exercises) to overcome them. Module 5 teaches subjects how to employ metacognitive therapy to overcome ruminating thoughts. Finally, in module 6, participants reflect on strategies they tried during their training and their potential future application. In every session, participants are invited to plan recreational activities and to incorporate them into their daily life, as per the behavioral activation approach. In addition, at the end of each session, participants are asked to select and complete at least one exercise during the following week. The training included interactive exercises, audio/video files, and downloadable material and was presented on a secured Web-based platform. As training was based on self-help, participants received only technical support via email. Screenshots of the intervention are available in the Multimedia Appendix 1.

Primary Outcome Measure

The study's primary outcome was insomnia severity, measured using the German version of the Insomnia Severity Index (ISI; [49]). This questionnaire consists of 7 items, each answered on a 5-point Likert scale, with responses ranging from 0 to 4 (total range 0-28). Summation scores are categorized as follows [52]: absence of insomnia (0-7), subthreshold insomnia (8-14), moderate insomnia (15-21), and severe insomnia (22-28). The instrument is a widely recommended outcome measure for clinical studies on insomnia symptoms [53] and has been validated as a Web-based measure [54]. The ISI has demonstrated high internal consistency in both community and clinical samples, with an internal consistency between 0.90 and 0.91 [52].

Secondary Outcome Measures

Secondary outcome measures assessed mental and work-related health and cognitive activities. Mental health outcomes included level of depression (Center for Epidemiological

Studies-Depression Scale, consisting of 20 items with ratings ranging from 0 to 3; total range 0-60; alpha=.88; [55]) and recuperation in sleep (Recuperation in Sleep subscale, consisting of 8 items with ratings ranging from 1 to 5; alpha=.85; [56]). Work-related health outcomes included the frequency of recreational activities after work over the past week (Recreation Experience and Activity Questionnaire, consisting of 21 items with ratings ranging from 0 to 4; total range 0-84; alpha=.77; [57]) and work ability (single-item score from the Work Ability Index; range 0-10; [58]). To assess subjects' self-rated number of full days on sick leave (absenteeism) and self-rated number of full days with reduced efficiency at work while feeling ill (presenteeism) over the past 3 months, the German Version of the Trimbos/Institute of Medical Technology Assessment questionnaire for costs associated with psychiatric illness was used [59], both at baseline and at 6-month follow-up. Cognitive activity was measured in 2 different ways: as work-related rumination (Cognitive Irritation subscale; 3 items ranging from 1 to 7; total range 3-28; alpha=.86; [60]) and as the subject's general tendency to worry (Penn State Worry Questionnaire, Ultra Brief Version, past week; 3 items ranging from 0 to 6; total range 0-18; alpha=.85; [61]).

Additional Measurements

Additional data collected included demographic variables and clients' self-rated level of satisfaction with the intervention (Client Satisfaction Questionnaire, adapted to the online context; 8 items ranging from 0 to 4; total range 0-32, alpha=.93; [62,63]).

Statistical Analyses

All analyses are reported in compliance with the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth guidelines for improving and standardizing the report of Web-based and mobile health interventions [64] using intention-to-treat (ITT) procedures. Analyses were performed using IBM SPSS version 25 by a researcher who was not otherwise involved in the developing and conducting of the study. No interim analyses for intervention efficacy were conducted. Reported *P* values were 2-sided, with the a priori threshold for statistical significance set at .05.

Missing Data

All participants completed the baseline assessment. At 8 weeks (T2), 25% (n=33 in the iCBT-I group and n=11 in the control group) of the data were missing; at T3, 29% (n=42 and n=9 in the iCBT-I group and the control group, respectively) of the data were missing. Assumptions of normality were assessed graphically using histograms. The robustness of the assumption regarding missing outcome data was examined in a series of sensitivity analyses (missing data patterns and inclusion of sample characteristics associated with having missing outcomes, eg, number of completed modules). All existing data of the primary and secondary outcomes, the number of completed modules, the grouping variable, and the interaction term as a product of group allocation with baseline scores of insomnia were used in the imputation model [65]. Multiple imputations with 100 estimates per missing value were conducted to handle missing data [66]. We used predictive mean matching (PMM) to reduce the possible bias introduced in a dataset through imputation by drawing real values sampled from the observed data. The PMM method ensures that imputed values are plausible and is robust if the normality assumption is violated [66]. The iterative Markov Chain Monte Carlo method was used, as it is appropriate when the data have an arbitrary (monotone or nonmonotone) missing pattern [66].

Efficacy of Training

Analysis was performed in the ITT population, including all randomized patients, using a linear mixed effects model to account for the repeated measures at baseline, postintervention, and at 6-month follow-up. Fixed effects included group allocation, time (preintervention, postintervention, and 6-month follow-up), and the group-by-time interaction. Random effects were run to account for between-subject variation. Cohen *d* and corresponding 95% CIs were computed for T2 and T3, comparing the means and standard deviations for both the iCBT-I and control groups immediately postintervention and at 6-month follow-up. For this purpose, the standard deviation was calculated from the standard error [67].

Meaningful Improvement and Symptom-Free Status

To detect a meaningful improvement in insomnia severity from T1 to T2 and from T1 to T3, proposed change scores from a study by Morin et al [52] were obtained. On the basis of this, ISI score differences of more than 4.6 points—between T1 and either T2 or T3—were considered a slight improvement. This change score is approximately as high as the reliable change index of 5.01 points in the ISI used in previous studies on the present iCBT-I [27,28]. According to Morin et al [52], in an individual subject, change scores of more than 8.4 points were used to find a moderate improvement and change scores of more than 9.9 points were used to find a marked improvement. To detect potential negative effects of the intervention, the number of participants with meaningful symptom deterioration, according to the change scores above, was assessed. To assess symptom-free status, an ISI score below 8 was considered symptom free [52]. In addition, the number needed to treat (NNT) was calculated by comparing the 2 groups for the number of participants (1) with or without a meaningful improvement in insomnia severity and (2) who became or failed to become symptom free.

Mediators of Interventional Effects

To examine both forms of cognitive activity, measured as worry and work-related rumination, as potential mediators of interventional effects on the primary outcome at T2, parallel multiple mediation analysis was conducted using the PROCESS software for SPSS (model 4), with bias-corrected bootstraps based on 10,000 bootstrap samples [68]. To establish temporal precedence, the postintervention scores of the mediators and 6-month follow-up scores for the primary outcome were used. Following the recommendation of Valente and McKinnon [69], baseline scores for the mediating variables and primary outcome were included as covariates in the model. Statistical significance of the mediation was achieved if the estimated 95% CI for the indirect effect did not overlap zero [68].

Moderators of Interventional Effects

The study sample was heterogeneous with respect to insomnia severity at baseline, as there were no inclusion criteria regarding the severity of the latter. To assess whether the intervention's effect on insomnia severity at postintervention and at 6-month follow-up is moderated by different levels of insomnia severity at baseline (before receiving the iCBT-I), simple moderation analysis was conducted using the PROCESS software for SPSS (model 1). For significant moderation, the Johnson-Neyman (J-N) procedure was employed [68,70] to identify the specific values of the moderator at which the groups differed significantly on the primary outcome at both assessment points.

Results

Participants

Figure 1 shows the flow of participants. In total, 177 individuals were randomized to iCBT-I (n=88) and control (n=89) groups. As the 10-month funding period that we were granted was inadequate to recruit the originally intended sample of 200 subjects and further recruitment was impossible, the final sample size was 177. With 177 subjects, the trial had 80% power to detect an intervention effect of d=0.42 at postintervention.

Baseline Characteristics

Table 1 summarizes the baseline characteristics of study participants. The sample consisted of 177 workers, of whom 116 were female (116/177, 65.5%), and the average age was 46.4 years (SD 9.8). Most participants (141/177, 79.7%) were employed full time and had an average occupational experience of 20.7 years (SD 9.7). More than half of the study's participants (103/177, 58.2%) had no previous experience with psychotherapy. Only 20 (20/177, 11.3%) participants had previously received psychotherapy for sleep problems, and just 33 (33/177, 18.6%) participants had undergone previous occupational health training. Most participants (123/177, 69.5%) reported clinically relevant insomnia symptoms (ISI score 15-28), whereas one-third of participants (54/177; 30.5%) had less or no insomnia symptoms (ISI score 0-14).

Table 1. Baseline characteristics.

Characteristics	Total (N=177)	Internet-delivered cognitive behavioral therapy for insomnia group (n=88)	Control group (n=89)
Sociodemographics, n (%)			
Females	116 (65.5)	59 (67)	57 (64)
Married/partnership	111 (62.7)	54 (61)	57 (64)
Age (years), mean (SD)	46.5 (9.8)	46.1 (9.5)	46.7 (9.7)
Working characteristics			
Years of occupational experience, mean (SD)	20.7 (9.7)	20.4 (9.8)	21.1 (10.5)
Permanent employment, n (%)	117 (66.1)	62 (71)	55 (62)
Employed fulltime, n (%)	141 (79.7)	69 (78)	72 (81)
Working sector, n (%)			
Health	35 (19.8)	15 (17)	20 (23)
Economy	34 (19.2)	21 (24)	13 (15)
Service	32 (18.1)	14 (16)	18 (20)
Social	27 (15.3)	14 (16)	13 (15)
Others	49 (27.7)	24 (27)	25 (28)
Experiences with training or psychotherapy, n (%)			
Occupational mental health training	33 (18.6)	18 (21)	15 (17)
Psychotherapy	74 (41.8)	34 (39)	40 (45)
Psychotherapy for sleeping problems	20 (11.3)	7 (8)	13 (15)
Insomnia severity, n (%)			
Severe (ISI[a] score 22-28)	17 (9.6)	6 (7)	11 (12)
Moderate (ISI score 15-21)	106 (59.9)	56 (64)	50 (56)
Subthreshold (ISI score 8-14)	49 (27.7)	24 (27)	25 (28)
Symptom free (ISI score 0-7)	5 (2.8)	2 (2)	3 (3)

[a]ISI: Insomnia Severity Index.

Intervention Use and User Satisfaction

Of the 88 individuals who were assigned to the iCBT-I group, 17 (19%) dropped out before completing module 1 of the intervention. Module 1 was completed by 71 (71/88, 81%) participants, module 2 was completed by 56 (56/88, 64%) participants, module 3 was completed by 51 (51/88, 58%) participants, module 4 was completed by 46 (46/88, 52%) participants, module 5 was completed by 39 (39/88, 44%) participants, and all modules were completed by 35 (35/88, 40%) participants. On average, 3.4 modules (SD 2.3) were completed in the training group. Those who completed the training needed, on average, 73.8 days (SD 55.6). With regard to intervention completer, the total completion time of the entire training was, on average, 5.0 hours (SD 1.3). For all modules, most participants required, on average, 0.5 to 1 hour.

Only 3 participants reported their reason for dropout; stated reasons were *technical problems*, *lack of time*, and *already sufficient help before the last module*. Participants' overall satisfaction with the training was high (n=54; mean 27.1 [SD 4.2]), with 93% (50/54) of the participants *satisfied in an overall, universal sense* (item 7, answering *generally yes* or *yes, completely*).

Primary Outcome Analysis—Insomnia Severity

For the primary outcome, the overall test of group-by-time interaction was significant ($P<.001$), with the intervention group reducing insomnia severity over time (Figure 2). Relative to controls, those who received the iCBT-I reported significantly reduced insomnia severity at T2 with a large effect size of Cohen $d=0.97$ and with a moderate-to-large effect size at 6-month follow-up (Cohen $d=0.86$). Table 2 displays all means, standard errors, and mean differences between groups for all outcome measures at T2 and T3 separately.

Figure 2. Comparison of internet-delivered cognitive behavioral therapy for insomnia and control groups on development of insomnia severity from baseline to 8 weeks after the training began and from baseline to 6-month follow-up. iCBT-I: internet-delivered cognitive behavioral therapy for insomnia.

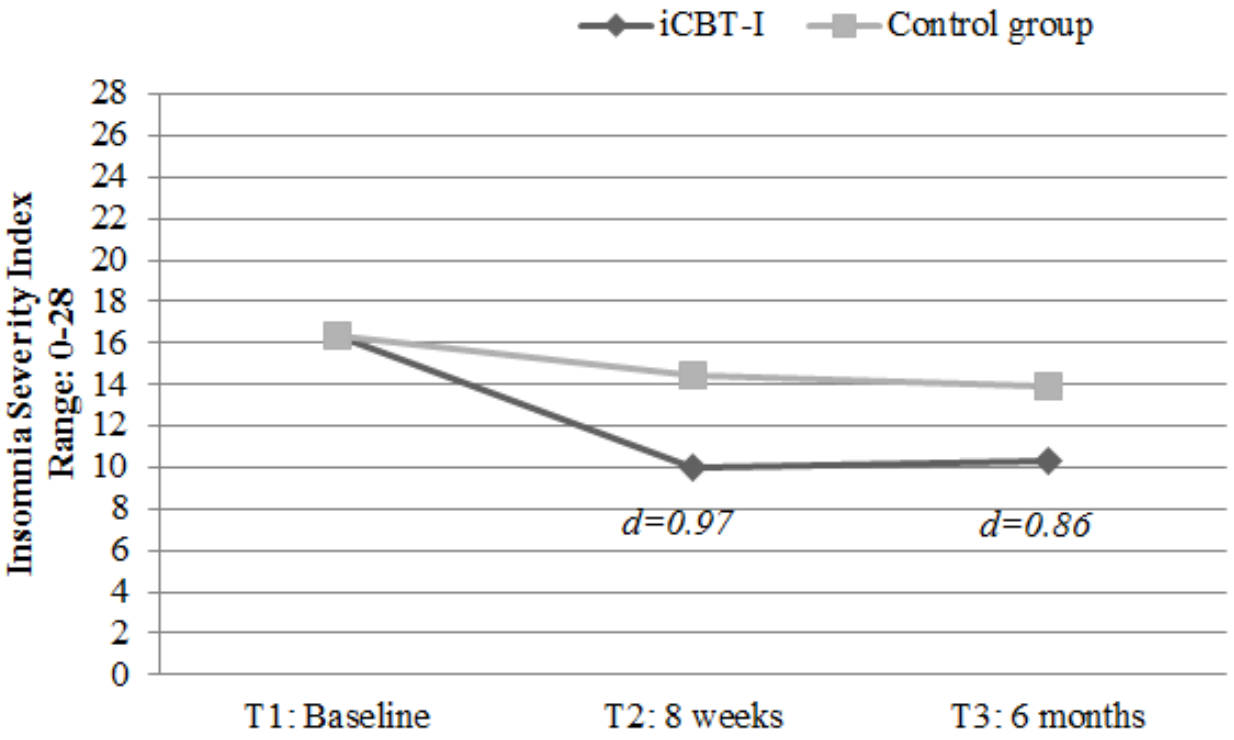


Table 2. Effects of internet-delivered cognitive behavioral therapy for insomnia group compared with the control group on primary and secondary outcomes.

Outcome	Internet-delivered cognitive behavioral therapy for insomnia group, mean (SE)	Control group, mean (SE)	Mean difference between groups (95% CI)	*P* value	Cohen *d* (95% CI)	Group × time, *P* value
Primary outcome						
Insomnia Severity Index						**<.001**
T1[a]	16.35 (0.46)	16.32 (0.46)	N/A[b]	N/A	N/A	
T2[c]	10.03 (0.48)	14.40 (0.48)	−4.36 (−5.69 to −3.03)	<.001	0.97 (0.66 to 1.28)	
T3[d]	10.30 (0.45)	13.93 (0.45)	−3.64 (−4.89 to −2.39)	<.001	0.86 (0.55 to 1.17)	
Mental health						
Center for Epidemiological Studies-Depression Scale						**<.001**
T1	22.08 (0.89)	20.78 (0.88)	N/A	N/A	N/A	
T2	15.40 (0.88)	17.91 (0.88)	−2.52 (−4.97 to −0.06)	.044	0.30 (0.01 to 0.60)	
T3	16.75 (0.83)	20.30 (0.82)	−3.55 (−5.86 to −1.24)	.003	0.46 (0.16 to 0.76)	
Recuperation in Sleep subscale[e]						**<.001**
T1	2.35 (0.06)	2.43 (0.06)	N/A	N/A	N/A	
T2	3.01 (0.07)	2.50 (0.07)	0.51 (0.31 to 0.70)	<.001	0.78 (0.47 to 1.08)	
T3	3.11 (0.06)	2.58 (0.06)	0.53 (0.35 to 0.71)	<.001	0.94 (0.63 to 1.25)	
Work-related health						
Recreation Experience and Activity Questionnaire[e]						**.121**
T1	47.07 (1.02)	45.11 (1.01)	N/A	N/A	N/A	
T2	51.4 (1.05)	46.65 (1.04)	4.75 (1.83 to 7.68)	.002	0.48 (0.18 to 078)	
T3	51.10 (1.00)	47.12 (0.99)	3.98 (1.21 to 6.75)	.005	0.43 (0.13 to 0.72)	
Work Ability Index[e]						**.006**
T1	6.88 (0.18)	6.94 (0.17)	N/A	N/A	N/A	
T3	7.14 (0.20)	6.46 (0.20)	0.68 (0.12 to 1.22)	.017	0.51 (0.21 to 0.81)	
Absenteeism[f]						**.869**
T1	4.62 (0.95)	2.69 (0.94)	N/A	N/A	N/A	
T3	4.71 (1.14)	3.14 (1.14)	1.57 (−1.61 to 4.75)	.333	0.21 (−0.09 to 0.50)	
Presenteeism[f]						**.103**
T1	10.69 (1.43)	13.17 (1.42)	N/A	N/A	N/A	
T3	5.94 (1.18)	12.39 (1.17)	−6.455 (−9.73 to 3.18)	<.001	0.83 (0.52 to 1.13)	
Cognitive activity						
Cognitive irritation subscale						**<.001**
T1	15.52 (0.37)	15.79 (0.36)	N/A	N/A	N/A	
T2	11.18 (0.40)	14.43 (0.39)	−3.25 (−4.35 to −2.15)	<.001	0.87 (0.57 to 1.18)	

Outcome	Internet-delivered cognitive behavioral therapy for insomnia group, mean (SE)	Control group, mean (SE)	Mean difference between groups (95% CI)	*P* value	Cohen *d* (95% CI)	Group × time, *P* value
T3	12.20 (0.39)	15.19 (0.39)	−2.99 (−4.08 to −1.90)	<.001	0.81 (0.51 to 1.12)	
Penn State Worry Questionnaire Past Week						**<.001**
T1	9.21 (0.41)	8.23 (0.41)	N/A	N/A	N/A	
T2	5.77 (0.40)	6.85 (0.39)	−1.08 (−2.18 to 0.02)	.05	0.29 (−0.01 to 0.59)	
T3	6.9 (0.39)	8.35 (0.39)	−1.46 (−2.55 to −0.36)	.009	0.39 (0.10 to 0.69)	

[a]At baseline.

[b]Not applicable.

[c]Postintervention (8 weeks postrandomization).

[d]At 6-month follow-up (4 months postintervention).

[e]Higher scores indicate better outcome.

[f]In relation to the previous 3 months.

Meaningful Improvement and Symptom-Free Status

Meaningful improvements are divided into 3 categories according to Morin et al [52]. A change score of 4.6 points in ISI was classified as a slight improvement, a change score of 8.4 points was classified as a moderate improvement, and a change score of 9.9 points was classified as a marked improvement. Meaningful improvements in insomnia severity and corresponding numbers needed to treat for T2 and T3 are shown in Table 3.

At T2, more participants in the iCBT-I group (57/88, 65%) reported a slight meaningful improvement in insomnia severity than the participants in the control group (16/88, 18%), which yielded an NNT of 2.14 (95% CI 1.68 to 2.94) to achieve a meaningful improvement in insomnia severity (Δ4.6) from baseline to postintervention. In addition, at T3, more participants in the iCBT-I group reported a meaningful improvement in insomnia severity (60/88, 68%) than those in the control group (25/88, 28%), corresponding to an NNT of 2.49 (95% CI 1.87 to 3.76).

At T2, 1 iCBT-I group member and 2 control subjects experienced meaningful deterioration (T2 score of >4.6 above the T1 score). At 6-month follow-up, the corresponding numbers were 1 and 2. For the other 2 categories (moderate and marked), no deteriorations could be detected for T2 and T3.

Significantly, more iCBT-I group participants (20/88, 23%) were symptom free at T2 than those in the control group (8/88, 9%), corresponding to an NNT of 7.28 (95% CI 4.11 to 31.68). The same was true at T3, with significantly more iCBT-I group participants (16/88, 18%) than control group participants (6/88, 7%) claiming to be symptom free, generating an NNT of 8.74 (95% CI 4.75 to 54.21).

Table 3. Meaningful improvements in insomnia severity from baseline to 8 weeks after the training began and from baseline to 6-month follow-up.

Meaningful improvements in insomnia severity	Meaningful improvement		Number needed to treat (95% CI)
	Internet-delivered cognitive behavioral therapy for insomnia group, n (%)	Control group, n (%)	
From baseline to 8 weeks after the training began			
Slight	57 (65)	16 (18)	2.14 (1.68 to 2.94)
Moderate	28 (32)	4 (4)	3.66 (2.63 to 5.99)
Marked	20 (23)	3 (3)	5.17(3.46 to 10.17)
From baseline to 6-month follow-up			
Slight	60 (68)	25 (28)	2.49 (1.87 to 3.76)
Moderate	16 (18)	6 (7)	8.74 (4.75 to 54.21)
Marked	75 (85)	86 (97)	8.87 (5.07 to 32.32)

Secondary Outcome Analyses

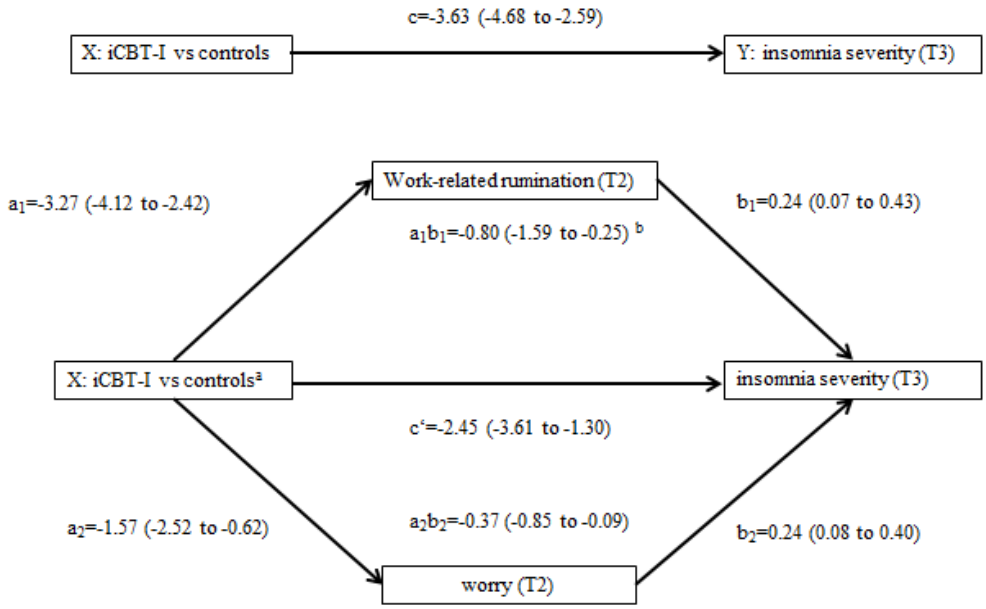
As Table 2 shows, significant differences in favor of the iCBT-I group were evident at both assessment points for the mental health outcomes of depression and recuperation in sleep. Work-related health outcome was found to significantly differ between the 2 groups, with regard to recreational activities, presenteeism, and work ability at T2 and T3. However, the between-group difference of absenteeism was nonsignificant at T3 (*P*=.33). Relating to cognitive activity, significant differences

between the 2 groups were identified for both measures of cognitive activity—work-related rumination and worry— at both assessment points. Effect sizes ranged from Cohen *d*=0.29 at T2 and Cohen *d*=0.39 at T3 for worry to Cohen *d*=0.87 at T2 and Cohen *d*=0.81 at T3 for work-related rumination.

Mediators of Interventional Effects

As depicted in Figure 3, the results of parallel multiple mediation analysis indicated that both forms of cognitive activity—work-related rumination (indirect effect: a_1b_1=−0.80; SE 0.34; 95% boot CI −1.59 to −0.25) and worry (indirect effect: a_2b_2=−0.37; SE 0.19; 95% boot CI −0.85 to −0.09)—significantly mediated the effect of the intervention on insomnia severity at 6-month follow-up. The direct effect of the intervention to reduce insomnia remained significant after incorporating the 2 mediators in the model: (direct effect: *c'*=−2.45; SE 0.59; 95% CI −3.61 to −1.30). This indicates that the intervention also reduced insomnia, independent of its indirect effects on worry and work-related rumination.

Figure 3. Parallel multiple mediation model with 6-month follow-up insomnia severity scores as the outcome variable, postintervention work-related rumination and worry scores as mediators, and baseline values of mediators and outcome as covariates. Interventiona X is coded 0=control groups and 1=iCBT-I group. bUnstandardized beta coefficients are shown, with 95% (bootstrapped biased corrected) CIs in parentheses. iCBT-I: internet-delivered cognitive behavioral therapy for insomnia.



Moderators of Interventional Effects

Moderation analysis showed a significant group×baseline insomnia severity interaction (beta=−.68; SE 0.13; t_{173}=−5.37; *P*<.001; 95% CI −0.93 to −0.43) immediately postintervention, indicating that the intervention's effect on insomnia severity was moderated by the severity of insomnia at baseline. Probing this effect, using the J-N technique, revealed a baseline ISI score of 12.08 as the point of transition between a statistically nonsignificant and significant effect of iCBT-I on postintervention insomnia severity. This means that, at T2, a significant group difference was identified for participants scoring 12 or greater on the baseline insomnia severity score (80% of all participants). No significant moderation effect of baseline levels of insomnia severity was detected for the intervention's effect on insomnia severity at the 6-month follow-up assessment (beta=−.21; SE 0.12; t_{173}=−1.67; *P*=.10; 95% CI −0.45 to 0.04).

Discussion

Principal Findings

In this study, we examined the efficacy of a self-help Web-based recovery training to improve sleep among the general working population. Results showed evidence that iCBT-I adapted for workers could be effective in the working population up to 6 months after start of the intervention and even workers with less sleep problems at baseline benefitted from the training in the longer term. Furthermore, it has been demonstrated that in addition to general cognitive activity, work-related cognitive activity mediates the interventions effect on sleep.

So far, we are aware of only 2 other randomized controlled trials that have evaluated the efficacy of self-help iCBT-I in a general working population without insomnia severity as an inclusion criterion. These studies revealed moderate-to-large effects on sleep in the short term [29,30]. This study adds to the limited evidence for the efficacy of iCBT-I in the general working population, demonstrating that a self-help iCBT-I not only has short-term benefits with moderate-to-large effects but also has

sustained benefits for up to 6 months. On average, participants changed from moderate insomnia to subthreshold category and managed to sustain their improvement. Even after 6 months, the NNT for a marked improvement (which represents a change score of about 10 points or more in insomnia severity) is still 8.87. Similar numbers were found in a recent meta-analysis of self-help intervention for depression at posttreatment [21]. Accordingly, 9 individuals are needed to achieve a marked improvement after 6 months, which represents a considerable effect taking the low costs of self-help iCBT-I and the context of universal prevention into account. The intervention's moderate-to-large effects on insomnia severity are consistent with results from recent meta-analyses that mainly focused on iCBT-I in clinical samples or in indicated prevention setting [15-18]. The results were also comparable with moderate-to-large effects found for iCBT-I in samples of the working population with elevated insomnia symptoms as the inclusion criterion [24-26]. However, consistent with expectations, effects also were less than in previous studies evaluating the efficacy of this training when used as either a guided or self-guided version by a sample of teachers with elevated insomnia symptoms [27,28]. One possible explanation is that the training's effects are greater in indicated prevention than in universal prevention settings, in which all interested workers could participate [45]. This might be explained by the fact that participants with raised levels of insomnia severity may be more motivated to implement what they were taught during the program and, therefore, produce more immediate postintervention results [45]. Another explanation might be found in the training's self-guided mode of delivery, as guided internet-based mental health interventions have been observed to induce greater reductions in symptoms than those that are self-guided [71].

As this study was conducted in the general working population and all interested workers who were motivated to improve their sleep through an iCBT-I were included, the sample was more heterogeneous with regard to insomnia severity at baseline than those of previously published studies in the working population [15-18,24,26-28]. Approximately 30% of our study participants reported less or even no symptoms of insomnia (ISI ≤14), and all participants would have been excluded from previous studies that restricted recruitment to those with moderate-to-severe insomnia.

Moderation analysis revealed a significant moderation effect of baseline insomnia severity level postintervention. At 6-month follow-up, no moderation effect was identified, indicating that participants' level of insomnia severity at baseline did not influence the intervention's efficacy over the long term. These findings are important because they indicate that a substantial portion of the working population—those reporting clinical relevant insomnia with an ISI score of 12 or more—were benefitting from the intervention immediately after the training program ended and that the training was even effective for participants with less baseline insomnia severity scores in the long term. This could be interpreted as a support for the idea of prevention. Those workers who face only less severe sleep complaints at baseline might benefit from participating in the interventions in the longer term, as they might equip themselves with skills that will help them to prevent that less severe symptoms turn not into more severe symptoms over time. In turn, workers without access to the intervention with less severe insomnia at baseline could develop more severe insomnia over the long term [72]. Considering that untreated subclinical insomnia can progress into clinical insomnia [11,72], in turn, potentially adversely affecting long-term health [9-11], the currently tested training might be a useful and effective measure for the prevention of more severe insomnia. Furthermore, the participation rate of 30.5% workers with no or subthreshold insomnia severity yields initial insight into the training's potential reach for workers with less sleep complaints in routine occupational health care.

It is notable that the intervention also was effective at reducing depressive symptoms, underlining the close relationship between sleep and depression [8]. These effects were even greater in magnitude at 6-month follow-up. In previously reported studies evaluating CBT-I, similar effects have been observed in those with moderately severe depressive symptoms, indicating that improvements in mood continued after intervention [73,74]. Considering these results, iCBT-I appears to be beneficial for both sleep and depression [8,16].

Apart from its positive effects on sleep and depression, the training we tested also exhibited meaningful effects on work-related health outcomes. Participants in the intervention group changed their health behaviors, indicating more frequent participation in positive recreational activities after work when the program ended and maintaining their healthy behavior change through 6 months. Similar effects on recreational activities were detected in 2 recent studies in teachers [27,28].

We identified mixed effects for work productivity, however. On the one hand, no significant effect on absenteeism was identified. This might be explained by a floor effect, as absenteeism was generally very low in the present sample. On the other hand, a significant, moderate effect was identified for presenteeism. This is also consistent with previously published findings, which indicate that insomnia primarily affects presenteeism, while it affects absenteeism only to a smaller extent [27,28]. It is also supported by health economics research that has documented how the positive economic effects of iCBT-I are mainly attributable to its influence on presenteeism [20]. Furthermore, the intervention affected subjects' ratings of their actual ability to work. Participants in the intervention group reported significantly higher work ability than those in the control group. These results are notable as workers suffering from sleep complaints often show poorer work performance [31].

Cognitive models of insomnia highlight the important role that negatively toned presleep cognitive activity plays in the development and persistence of insomnia [32,37]. For this reason, the present intervention also focused on reducing presleep cognitive activity, thereby considering both general and work-related cognitive activities. Small effects were identified in terms of reducing general cognitive activity and measured as worry. Conversely, a reduction in work-related cognitive activity, in terms of work-related rumination, showed large effects. These effects are important because work-related

rumination has been shown to be a mediator between work-stress and impaired sleep [33], and both rumination [75] and insomnia [7] are predictors of future depression.

Despite the extensive literature citing cognitive activity as a key developmental and maintenance factor for insomnia [32,37], little is known about how iCBT-I improves sleep by reducing presleep cognitive activity [28,36,43,76]. Our secondary aim was to exploratory examine whether, in addition to the role of general cognitive activity, work-related cognitive activity might be a mediator of the intervention's efficacy. Results of parallel multiple mediation analyses help to clarify such mechanisms, revealing that our intervention's effect on sleep was mediated by reducing the following 2 forms of cognitive activity that we studied: worry and work-related rumination. In other words, in our sample, the training indirectly enhanced sleep by reducing general worry and work-related rumination. This suggests that tailoring an intervention to the specific needs of workers to reduce work-related cognitive activity can be a promising approach to effectively reduce sleep problems of workers. This further supports prior research documenting the benefits of tailoring interventions to users' life domains [77].

To summarize, our results suggest that reducing presleep cognitive activity could be an important component of any effective intervention for insomnia, and that including components that target reducing work-related cognitive activity might be an important contribution to designing interventions that reduce sleep complaints in workers.

Limitations

The following limitations of this trial must be acknowledged. First, because of the restricted funding period, the final sample size differed from the sample size initially intended. However, given the intervention's sizeable effect on the primary outcome, we believe that our conclusions are not substantially affected by the smaller sample size. Second, the intervention was not offered within a typical workplace setting (eg, providing the intervention within the company or informing about participation options within the confines of employee assemblies or other in-house communication). The recruitment strategy used in this study (ie, mass media and email distribution list from one of the largest health insurance companies in Germany) presents a complementary way of approaching workers who are interested in improving their sleep. Therefore, it is unclear if our results can be generalized to routine occupational health care settings approaching workers within the company. Nonetheless, the applied recruiting strategy here mimics 1 possibility of how internet-delivered mental health interventions are made available for workers in Germany.

Third, we used a waitlist-control and no attention control group. Thus, it is not possible to determine whether and to what extent the effects of the intervention were because of specific iCBT-I content or because of nonspecific support provided by participating in an intervention. Moreover, one might expect that the delayed access to the training for the control group leads to a sense of disappointment and rumination about sleep while waiting to receive the intervention. Measuring disappointment would have been a way of controlling such sources of bias, which should be considered in further studies with waitlist-control groups. Moreover, in future studies, it would be of practical importance to employ other self-help interventions as a comparator (ie, self-help books to determine the unique contribution of the internet-delivered format) [78]. However, it should be mentioned that previous studies largely used similar control groups; therefore, the results of this study can be compared with existing research [24,27-30].

Fourth, there was a high dropout rate as defined by the study protocol and by the completion of modules among participants in the intervention group. This could lead to an overestimation of the intervention effect because it is assumed that participants who were not satisfied with the intervention nor appeared to have experienced any benefit may not have answer any further questionnaires postintervention and at 6-month follow-up [79]. Thus, these participants may not have provided data. On the other hand, these participants could also be early completer who benefit from the intervention but do not complete the study protocol and finish the intervention early [80]. In addition, to reduce a biased estimate of the missing data, the number of modules completed was included in the imputation model [65].

Fifth, although no inclusion criteria with regard to insomnia severity were set up in context of universal prevention, approximately 60% of participants had moderate insomnia symptoms, and approximately 30% had subthreshold insomnia symptoms or were symptom free. As recruiting participants in the general working population for this study overlapped timewise with another study, it is likely that the percentage of workers with severe insomnia symptoms might be underestimated. This said, the results of this study are the first indication of the acceptance and efficacy of iCBT-I in a universal prevention setting in routine occupational health care and demonstrate that a substantial proportion of workers with no or less sleep problems are motivated to participate.

Finally, no measures were assessed to determine which of the training elements were the most effective and which elements were less effective. However, results from mediation analysis suggest that techniques that reduce presleep cognitive activity played an important role in the training's efficacy. To explore which of the intervention's components work best, further research is needed.

Conclusions

In conclusion, the results of this study give further indications that an internet-delivered self-help CBT-I adapted for workers has stable effects up to 6 months after the training began.

Moreover, the presented training was open to all workers and a substantial proportion with less sleep problems participated, who benefitted from the training in the longer term. In this way, the present findings strengthen the arguments as to the generalizability and robustness of prior results on this training's efficacy [27,28] and also that of prior results of iCBT-I in the working population [24-26,29,30]. The intervention's substantial effect is noteworthy, as the training was delivered in a self-help form that increases its potential for large-scale and cost-effective implementation in universal prevention. Conclusively, from a public health perspective, more workers who are potentially in need of an intervention to improve their sleep could be reached.

Acknowledgments

The European Union funded this study, project number: EFRE: CCI 2007DE161PR001. The authors thank Hanne Horvath, who was considerably involved in the development of the training and evaluation.

Authors' Contributions

DL and DDE contributed to the design of the study. DL led the development of the intervention content. DB performed the outcome analyses. DB drafted the first proof of the paper and integrated coauthor comments from DL, DDE, and KS and edits. All the authors contributed to the further writing of the paper and approved the final paper.

Conflicts of Interest

DL and DDE are stakeholders in the Institute for Online Health Training, which aims to transfer scientific knowledge related to this research into routine health care. DB and KS have no conflicts of interest to declare.

Multimedia Appendix 1
Screenshots of the intervention.
[PDF File (Adobe PDF File), 614 KB - jmir_v22i1e13346_app1.pdf]

Multimedia Appendix 2
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2543 KB - jmir_v22i1e13346_app2.pdf]

References

1. Morin CM, Drake CL, Harvey AG, Krystal AD, Manber R, Riemann D, et al. Insomnia disorder. Nat Rev Dis Primers 2015 Sep 3;1:15026. [doi: 10.1038/nrdp.2015.26] [Medline: 27189779]
2. Storm A. DAK- Health Report 2017: Analysis of data on work disability. Update: Sleep disorders. Contributions to health economics and health services research. Heidelberg: medhochzwei; 2017.
3. Sivertsen B, Overland S, Bjorvatn B, Maeland JG, Mykletun A. Does insomnia predict sick leave? The Hordaland Health Study. J Psychosom Res 2009 Jan;66(1):67-74. [doi: 10.1016/j.jpsychores.2008.06.011] [Medline: 19073296]
4. Sofi F, Cesari F, Casini A, Macchi C, Abbate R, Gensini GF. Insomnia and risk of cardiovascular disease: a meta-analysis. Eur J Prev Cardiol 2014 Jan;21(1):57-64. [doi: 10.1177/2047487312460020] [Medline: 22942213]
5. Li M, Zhang X, Hou W, Tang Z. Insomnia and risk of cardiovascular disease: a meta-analysis of cohort studies. Int J Cardiol 2014 Oct 20;176(3):1044-1047. [doi: 10.1016/j.ijcard.2014.07.284] [Medline: 25156850]
6. Patel SR, Hu FB. Short sleep duration and weight gain: a systematic review. Obesity (Silver Spring) 2008 Mar;16(3):643-653 [FREE Full text] [doi: 10.1038/oby.2007.118] [Medline: 18239586]
7. Pigeon WR, Bishop TM, Krueger KM. Insomnia as a precipitating factor in new onset mental illness: a systematic review of recent findings. Curr Psychiatry Rep 2017 Aug;19(8):44. [doi: 10.1007/s11920-017-0802-x] [Medline: 28616860]
8. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
9. Mallon L, Broman JE, Hetta J. Sleep complaints predict coronary artery disease mortality in males: a 12-year follow-up study of a middle-aged Swedish population. J Intern Med 2002 Mar;251(3):207-216 [FREE Full text] [doi: 10.1046/j.1365-2796.2002.00941.x] [Medline: 11886479]
10. National Institutes of Health. National Institutes of Health State of the Science Conference statement on Manifestations and Management of Chronic Insomnia in Adults, June 13-15, 2005. Sleep 2005 Sep;28(9):1049-1057. [doi: 10.1093/sleep/28.9.1049] [Medline: 16268373]
11. Suka M, Yoshida K, Sugimori H. Persistent insomnia is a predictor of hypertension in Japanese male workers. J Occup Health 2003 Nov;45(6):344-350 [FREE Full text] [doi: 10.1539/joh.45.344] [Medline: 14676413]
12. Riemann D, Baglioni C, Bassetti C, Bjorvatn B, Groselj LD, Ellis JG, et al. European guideline for the diagnosis and treatment of insomnia. J Sleep Res 2017 Dec;26(6):675-700 [FREE Full text] [doi: 10.1111/jsr.12594] [Medline: 28875581]
13. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
14. Ebert DD, van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet and mobile-based psychological interventions: applications, efficacy and potential for improving mental health. A report of the EFPA e-health taskforce. Eur Psychol 2018;23(2):167-187. [doi: 10.1027/1016-9040/a000318]

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. Ye Y, Zhang Y, Chen J, Liu J, Li X, Liu Y, et al. Internet-based cognitive behavioral therapy for insomnia (ICBT-i) improves comorbid anxiety and depression - A meta-analysis of randomized controlled trials. PLoS One 2015;10(11):e0142258 [FREE Full text] [doi: 10.1371/journal.pone.0142258] [Medline: 26581107]
17. Ye Y, Chen N, Chen J, Liu J, Lin L, Liu Y, et al. Internet-based cognitive-behavioural therapy for insomnia (ICBT-i): a meta-analysis of randomised controlled trials. BMJ Open 2016 Nov 30;6(11):e010707 [FREE Full text] [doi: 10.1136/bmjopen-2015-010707] [Medline: 27903557]
18. Cheng SK, Dizon J. Computerised cognitive behavioural therapy for insomnia: a systematic review and meta-analysis. Psychother Psychosom 2012;81(4):206-216. [doi: 10.1159/000335379] [Medline: 22585048]
19. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry 2016 Apr;3(4):333-341. [doi: 10.1016/S2215-0366(15)00536-2] [Medline: 26827250]
20. Thiart H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. Sleep 2016 Oct 1;39(10):1769-1778 [FREE Full text] [doi: 10.5665/sleep.6152] [Medline: 27450686]
21. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
22. Nakada Y, Sugimoto A, Kadotani H, Yamada N. Verification of effect of sleep health education program in workplace: a quasi-randomized controlled trial. Ind Health 2018 Feb 7;56(1):20-29 [FREE Full text] [doi: 10.2486/indhealth.2017-0022] [Medline: 28855438]
23. Schiller H, Söderström M, Lekander M, Rajaleid K, Kecklund G. A randomized controlled intervention of workplace-based group cognitive behavioral therapy for insomnia. Int Arch Occup Environ Health 2018 May;91(4):413-424 [FREE Full text] [doi: 10.1007/s00420-018-1291-x] [Medline: 29387936]
24. Bostock S, Luik AI, Espie CA. Sleep and productivity benefits of digital cognitive behavioral therapy for insomnia: a randomized controlled trial conducted in the workplace environment. J Occup Environ Med 2016 Jul;58(7):683-689. [doi: 10.1097/JOM.0000000000000778] [Medline: 27257747]
25. Espie CA, Pawlecki B, Waterfield D, Fitton K, Radocchia M, Luik AI. Insomnia symptoms and their association with workplace productivity: cross-sectional and pre-post intervention analyses from a large multinational manufacturing company. Sleep Health 2018 Jun;4(3):307-312. [doi: 10.1016/j.sleh.2018.03.003] [Medline: 29776626]
26. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 1;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
27. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015 Mar;41(2):164-174 [FREE Full text] [doi: 10.5271/sjweh.3478] [Medline: 25590336]
28. Ebert DD, Berking M, Thiart H, Riper H, Laferton JA, Cuijpers P, et al. Restoring depleted resources: efficacy and mechanisms of change of an internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015 Dec;34S:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
29. Barnes CM, Miller JA, Bostock S. Helping employees sleep well: effects of cognitive behavioral therapy for insomnia on work outcomes. J Appl Psychol 2017 Jan;102(1):104-113. [doi: 10.1037/apl0000154] [Medline: 27690480]
30. Suzuki E, Tsuchiya M, Hirokawa K, Taniguchi T, Mitsuhashi T, Kawakami N. Evaluation of an internet-based self-help program for better quality of sleep among Japanese workers: a randomized controlled trial. J Occup Health 2008;50(5):387-399 [FREE Full text] [doi: 10.1539/joh.l7154] [Medline: 18716392]
31. Litwiller B, Snyder LA, Taylor WD, Steele LM. The relationship between sleep and work: a meta-analysis. J Appl Psychol 2017 Apr;102(4):682-699. [doi: 10.1037/apl0000169] [Medline: 27893255]
32. Harvey AG. A cognitive model of insomnia. Behav Res Ther 2002 Aug;40(8):869-893. [doi: 10.1016/s0005-7967(01)00061-4] [Medline: 12186352]
33. Berset M, Elfering A, Lüthy S, Lüthi S, Semmer NK. Work stressors and impaired sleep: rumination as a mediator. Stress Health 2011 Apr;27(2):e71-e82. [doi: 10.1002/smi.1337] [Medline: 27486625]
34. Brosschot JF, Gerin W, Thayer JF. The perseverative cognition hypothesis: a review of worry, prolonged stress-related physiological activation, and health. J Psychosom Res 2006 Feb;60(2):113-124. [doi: 10.1016/j.jpsychores.2005.06.074] [Medline: 16439263]
35. Radstaak M, Geurts SA, Beckers DG, Brosschot JF, Kompier MA. Work stressors, perseverative cognition and objective sleep quality: a longitudinal study among Dutch Helicopter Emergency Medical Service (HEMS) Pilots. J Occup Health 2014;56(6):469-477 [FREE Full text] [doi: 10.1539/joh.14-0118-OA] [Medline: 25744087]

36. Gosling JA, Batterham P, Ritterband L, Glozier N, Thorndike F, Griffiths KM, et al. Online insomnia treatment and the reduction of anxiety symptoms as a secondary outcome in a randomised controlled trial: The role of cognitive-behavioural factors. Aust N Z J Psychiatry 2018 Dec;52(12):1183-1193. [doi: 10.1177/0004867418772338] [Medline: 29717621]
37. Riemann D, Spiegelhalder K, Feige B, Voderholzer U, Berger M, Perlis M, et al. The hyperarousal model of insomnia: a review of the concept and its evidence. Sleep Med Rev 2010 Feb;14(1):19-31. [doi: 10.1016/j.smrv.2009.04.002] [Medline: 19481481]
38. Querstret D, Cropley M. Assessing treatments used to reduce rumination and/or worry: a systematic review. Clin Psychol Rev 2013 Dec;33(8):996-1009. [doi: 10.1016/j.cpr.2013.08.004] [Medline: 24036088]
39. Brosschot JF, Van Dijk E, Thayer JF. Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period. Int J Psychophysiol 2007 Jan;63(1):39-47. [doi: 10.1016/j.ijpsycho.2006.07.016] [Medline: 17020787]
40. Pillai V, Roth T, Mullins HM, Drake CL. Moderators and mediators of the relationship between stress and insomnia: stressor chronicity, cognitive intrusion, and coping. Sleep 2014 Jul 1;37(7):1199-1208 [FREE Full text] [doi: 10.5665/sleep.3838] [Medline: 25061248]
41. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
42. Carney CE, Harris AL, Moss TG, Edinger JD. Distinguishing rumination from worry in clinical insomnia. Behav Res Ther 2010 Jun;48(6):540-546 [FREE Full text] [doi: 10.1016/j.brat.2010.03.004] [Medline: 20362977]
43. Luik AI, Kyle SD, Espie CA. Digital Cognitive Behavioral Therapy (dCBT) for Insomnia: a State-of-the-Science Review. Curr Sleep Med Rep 2017;3(2):48-56 [FREE Full text] [doi: 10.1007/s40675-017-0065-4] [Medline: 28553574]
44. Schwartz DR, Carney CE. Mediators of cognitive-behavioral therapy for insomnia: a review of randomized controlled trials and secondary analysis studies. Clin Psychol Rev 2012 Nov;32(7):664-675. [doi: 10.1016/j.cpr.2012.06.006] [Medline: 22935473]
45. 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 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
46. Rose G. Sick individuals and sick populations. Int J Epidemiol 1985 Mar;14(1):32-38. [doi: 10.1093/ije/14.1.32] [Medline: 3872850]
47. Beck AT, Steer RA, Brown GK. BDI-II, Beck depression inventory: manual. San Antonio, TX: Harcourt Brace; 1996.
48. Kreiner GE, Hollensbe EC, Sheep ML. Balancing borders and bridges: negotiating the work-home interface via boundary work tactics. Acad Manag J 2009;52(4):704-730. [doi: 10.5465/amj.2009.43669916]
49. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
50. Hewitt CE, Torgerson DJ. Is restricted randomisation necessary? Br Med J 2006 Jun 24;332(7556):1506-1508 [FREE Full text] [doi: 10.1136/bmj.332.7556.1506] [Medline: 16793819]
51. Morin CM. Cognitive behavioral therapy for chronic insomnia: state of the science versus current clinical practices. Ann Intern Med 2015 Aug 4;163(3):236-237. [doi: 10.7326/M15-1246] [Medline: 26052868]
52. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 1;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
53. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
54. Thorndike FP, Ritterband LM, Saylor DK, Magee JC, Gonder-Frederick LA, Morin CM. Validation of the insomnia severity index as a web-based measure. Behav Sleep Med 2011;9(4):216-223. [doi: 10.1080/15402002.2011.606766] [Medline: 22003975]
55. Hautzinger M, Bailer M, Hofmeister D, Keller F. Center for Epidemiological Studies Depression Scale. Goettingen: Hogrefe; 2012.
56. Görtelmeyer R. SF-A/R and SF-B/R: Sleep Questionnaires A and B - Revised version. Goettingen: Hogrefe; 2011.
57. Lehr D. The Recreation Experience and Activity Questionnaire (ReaQ). In: Koch S, Lehr D, Hillert A, editors. Burnout und chronischer beruflicher Stress. Goettingen: Hogrefe; 2015:107-108.
58. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health--a prospective study of women on long-term sick leave. Scand J Work Environ Health 2010 Sep;36(5):404-412 [FREE Full text] [doi: 10.5271/sjweh.2917] [Medline: 20372766]
59. Bouwmans C, de Jong K, Timman R, Zijlstra-Vlasveld M, van der Feltz-Cornelis C, Swan ST, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013 Jun 15;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
60. Mohr G, Rigotti T, Müller A. Irritation Scale for the Assessment of Psychological Strain in Work Contexts. Goettingen: Hogrefe; 2007.

61. Stöber J, Bittencourt J. Weekly assessment of worry: an adaptation of the Penn State Worry Questionnaire for monitoring changes during treatment. Behav Res Ther 1998 Jun;36(6):645-656. [doi: 10.1016/s0005-7967(98)00031-x] [Medline: 9648338]
62. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
63. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
64. 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]
65. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Br Med J 2009 Jun 29;338:b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
66. van Buuren S. Flexible Imputation of Missing Data. Second Edition. New York: CRC Press; 2018.
67. Higgins JP. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: Wiley-Blackwell; 2008.
68. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: The Guilford Press; 2018.
69. Valente MJ, MacKinnon DP. Comparing models of change to estimate the mediated effect in the pretest-posttest control group design. Struct Equ Modeling 2017;24(3):428-450 [FREE Full text] [doi: 10.1080/10705511.2016.1274657] [Medline: 28845097]
70. Hayes AF, Rockwood NJ. Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation. Behav Res Ther 2017 Nov;98:39-57. [doi: 10.1016/j.brat.2016.11.001] [Medline: 27865431]
71. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — A systematic review. Internet Interv 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
72. 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 9;169(5):447-453. [doi: 10.1001/archinternmed.2008.610] [Medline: 19273774]
73. Taylor DJ, Lichstein KL, Weinstock J, Sanford S, Temple JR. A pilot study of cognitive-behavioral therapy of insomnia in people with mild depression. Behav Ther 2007 Mar;38(1):49-57. [doi: 10.1016/j.beth.2006.04.002] [Medline: 17292694]
74. Swift N, Stewart R, Andiappan M, Smith A, Espie CA, Brown JS. The effectiveness of community day-long CBT-I workshops for participants with insomnia symptoms: a randomised controlled trial. J Sleep Res 2012 Jun;21(3):270-280 [FREE Full text] [doi: 10.1111/j.1365-2869.2011.00940.x] [Medline: 21848573]
75. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008 Sep;3(5):400-424. [doi: 10.1111/j.1745-6924.2008.00088.x] [Medline: 26158958]
76. Lancee J, Effting M, van der Zweerde T, van Daal L, van Straten A, Kamphuis JH. Cognitive processes mediate the effects of insomnia treatment: evidence from a randomized wait-list controlled trial. Sleep Med 2019 Feb;54:86-93. [doi: 10.1016/j.sleep.2018.09.029] [Medline: 30529782]
77. Blonk RW, Brenninkmeijer V, Lagerveld SE, Houtman IL. Return to work: a comparison of two cognitive behavioural interventions in cases of work-related psychological complaints among the self-employed. Work Stress 2006;20(2):129-144. [doi: 10.1080/02678370600856615]
78. van Straten A, Cuijpers P. Self-help therapy for insomnia: a meta-analysis. Sleep Med Rev 2009 Feb;13(1):61-71. [doi: 10.1016/j.smrv.2008.04.006] [Medline: 18952469]
79. 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]
80. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

Abbreviations

CBT-I: cognitive behavioral therapy for insomnia
DRKS: German Clinical Trials Register
iCBT-I: internet-delivered cognitive behavioral therapy for insomnia
ISI: Insomnia Severity Index
ITT: intention-to-treat
J-N: Johnson-Neyman
NNT: number needed to treat

PMM: predictive mean matching

Edited by G Eysenbach; submitted 09.01.19; peer-reviewed by A Luik, P Saskin, A Farmer, P Batterham; comments to author 27.04.19; revised version received 22.09.19; accepted 26.09.19; published 07.01.20.

Please cite as:
Behrendt D, Ebert DD, Spiegelhalder K, Lehr D
Efficacy of a Self-Help Web-Based Recovery Training in Improving Sleep in Workers: Randomized Controlled Trial in the General Working Population
J Med Internet Res 2020;22(1):e13346
URL: https://www.jmir.org/2020/1/e13346
doi:10.2196/13346
PMID:31909725

Original Paper

Self-Care Monitoring of Heart Failure Symptoms and Lung Impedance at Home Following Hospital Discharge: Longitudinal Study

Ina Thon Aamodt[1,2], RN, MNSc; Edita Lycholip[3,4], RN, PhD; Jelena Celutkiene[3,4], MD, PhD; Thomas von Lueder[5], MD, PhD; Dan Atar[6,7], MD, PhD; Ragnhild Sørum Falk[8], PhD; Ragnhild Hellesø[2], RN, PhD; Tiny Jaarsma[9], RN, PhD; Anna Strömberg[10,11*], RN, PhD; Irene Lie[1*], RN, PhD

[1]Centre for Patient-Centered Heart and Lung Research, Department of Cardiothoracic Surgery, Oslo University Hospital, Ullevål, Oslo, Norway
[2]Department of Nursing Science, Institute of Health and Society, University of Oslo, Oslo, Norway
[3]Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine of the Faculty of Medicine, Vilnius University, Vilnius, Lithuania
[4]Center of Cardiology and Angiology, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania
[5]Department of Cardiology B, Oslo University Hospital, Ullevål, Oslo, Norway
[6]Department of Cardiology, Oslo University Hospital, Ullevål, Oslo, Norway
[7]Institute of Clinical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway
[8]Research Support Services, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
[9]Division of Nursing, Department of Social and Welfare Studies, Linköping University, Norrköping, Sweden
[10]Division of Nursing, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
[11]Department of Cardiology, Linköping University, Linköping, Sweden
[*]these authors contributed equally

Corresponding Author:
Ina Thon Aamodt, RN, MNSc
Centre for Patient-Centered Heart and Lung Research, Department of Cardiothoracic Surgery
Oslo University Hospital
Ullevål
Building 63, Box 4956 Nydalen
Oslo
Norway
Phone: 47 48090883
Email: inamarieaamodt@gmail.com

Abstract

Background: Self-care is key to the daily management of chronic heart failure (HF). After discharge from hospital, patients may struggle to recognize and respond to worsening HF symptoms. Failure to monitor and respond to HF symptoms may lead to unnecessary hospitalizations.

Objective: This study aimed to (1) determine the feasibility of lung impedance measurements and a symptom diary to monitor HF symptoms daily at home for 30 days following hospital discharge and (2) determine daily changes in HF symptoms of pulmonary edema, lung impedance measurements, and if self-care behavior improves over time when patients use these self-care monitoring tools.

Methods: This study used a prospective longitudinal design including patients from cardiology wards in 2 university hospitals—one in Norway and one in Lithuania. Data on HF symptoms and pulmonary edema were collected from 10 participants (mean age 64.5 years; 90% (9/10) male) with severe HF (New York Heart Association classes III and IV) who were discharged home after being hospitalized for an HF condition. HF symptoms were self-reported using the Memorial Symptom Assessment Scale for Heart Failure. Pulmonary edema was measured by participants using a noninvasive lung impedance monitor, the CardioSet Edema Guard Monitor. Informal caregivers aided the participants with the noninvasive measurements.

Results: The prevalence and burden of shortness of breath varied from participants experiencing them daily to never, whereas lung impedance measurements varied for individual participants and the group participants, as a whole. Self-care behavior score

improved significantly (*P*=.007) from a median of 56 (IQR range 22-75) at discharge to a median of 81 (IQR range 72-98) 30 days later.

Conclusions: Noninvasive measurement of lung impedance daily and the use of a symptom diary were feasible at home for 30 days in HF patients. Self-care behavior significantly improved after 30 days of using a symptom diary and measuring lung impedance at home. Further research is needed to determine if daily self-care monitoring of HF signs and symptoms, combined with daily lung impedance measurements, may reduce hospital readmissions.

(J Med Internet Res 2020;22(1):e15445) doi:10.2196/15445

KEYWORDS

heart failure; telemedicine; lung impedance; diary; self-care; prospective study

Introduction

Background

Self-care is recognized as an important aspect of daily management of heart failure (HF) and recommended in international guidelines [1-3]. Self-care is defined as a process to maintain health through health-promoting practices and managing illness [4]. Self-care has a positive effect on HF prognosis, HF readmission rates, functional capacity, and well-being [1,4-6]. An important aspect of self-care is self-monitoring, which requires patients to observe and recognize changes in symptoms and signs of HF [4]. Many HF patients struggle with recognizing, interpreting, and taking appropriate action if worsening of symptoms occurs [7,8]. Fluid retention (congestion) often develops gradually in HF, and some HF patients only experience few or atypical symptoms [1]. Improved knowledge among HF patients on possible symptoms of fluid retention and knowing when to consult health care professionals may prevent hospitalizations [9]. Moreover, 1 out of 5 older HF patients are readmitted within 30 days after discharge [10]. In these readmitted patients, HF prognosis remains poor [11,12]. Reducing the readmission rate during the first 30 days after hospital discharge is important to decrease the disease burden experienced by patients and informal caregivers and the economic burden on the health care system [11,13].

The 5 most characteristic HF symptoms of fluid retention are shortness of breath, shortness of breath when supine, shortness of breath that awakens the patient during sleep, feeling tired, and ankle swelling [1]. Moreover, comorbidity is common in patients diagnosed with HF, and some comorbid conditions may have similar symptoms as those of fluid retention, for example, shortness of breath caused by chronic obstructive lung disease [14]. HF patients might therefore benefit from tools that aid them in monitoring and managing their postdischarge symptoms, for example, a diary or a fluid monitoring device. Using a symptom diary, HF patients can successfully detect small changes in symptoms [15,16]. Use of a symptom diary has been associated with improved self-care, survival, and quality of life and fewer HF-related hospitalizations [15-18]. Symptom diaries contain instructions for HF patients and a list of various self-reported HF symptoms in addition to space where patients can provide their weight, blood pressure, heart rate, and comments.

Another promising method for self-monitoring is using noninvasive lung impedance devices to measure pulmonary congestion before HF symptoms are recognized by the patient [19,20]. As pulmonary edema or congestion develops, lung impedance decreases.

Objectives

Knowledge is lacking for self-care approaches that combine the use of a self-reported symptom diary with noninvasive lung impedance measurements. The aim of this study was, therefore, to assess patients' HF symptoms, lung impedance, and self-care behavior at home for 30 days after hospital discharge. This study sought answers to the following research questions:

1. How feasible is it for patients with HF and their caregivers to measure symptoms and signs daily at home using a diary and a noninvasive device for measuring lung impedance during a 30-day period after discharge from hospital?

2. How do daily HF symptoms and lung impedance change during the 30-day assessment period?

3. How does self-care behavior change when patients use a symptom diary and a noninvasive device to measure lung impedance during the assessment period?

Methods

Study Design and Setting

This longitudinal observational design study was conducted from May 2017 to November 2017 using eligible patients discharged from cardiology wards in 2 university hospitals in Norway and Lithuania. Norway is a high-income country ranked as number 28 by the World Bank, and Lithuania is ranked as number 84 [21]. The health care system in Lithuania is a mixed system funded by the National Health Insurance Fund and the state, whereas funding in Norway is provided by public sources [22,23]. The prevalence of HF was 3.09% and 1.71% in Lithuania [24] and Norway [25], respectively.

Inclusion Criteria

Patients were eligible if they were hospitalized with a primary diagnosis of HF, aged older than 18 years, fluent in Norwegian or Lithuanian, and possessed sufficient cognitive abilities to understand and complete the study protocol. Cognitive abilities were judged by the nurses or cardiologists at the hospital ward.

Only patients with New York Heart Association (NYHA) Functional Classification III and IV were included in the study.

Exclusion Criteria

Patients with severe HF in need of surgical intervention, advanced chronic kidney disease defined as estimated glomerular filtration rate less than 25 mL per min per 1.73 m^2 [20], a body weight of greater than 150 kg, documented major depression, or short expected survival time were excluded.

Data Collection and Instruments

The symptom diary comprised 3 components (Table 1) and was self-administered daily at home for 30 days after hospital discharge. Self-reported questionnaires and a clinical examination were administered both at discharge and at the outpatient clinic 30 days later after completion of the at-home data collection period (see Table 1).

Table 1. Data collection instruments administered at discharge, at home or at the outpatient clinic.

Data collection instruments	Discharge	Home for 30 days	Outpatient clinic
Symptom diary			
Memorial Symptom Assessment Scale-Heart Failure	—[a]	x[b]	—
Lung impedance	—	x	—
Medication on demand	—	x	—
Questionnaire			
European Heart Failure Self-care Behavior Scale	x	—	x
Clinical examination	x	—	x

[a]Not applicable.

[b]The x indicates the time and place of data collection.

Procedure

Candidate participants were identified by cardiologists at the hospital wards. Before discharge, the included participants were instructed on how to use the printed symptom diary, which consisted of spaces to rate symptoms and to write down lung impedance measurements and on-demand medications. Participants and informal caregivers were trained at their home on how to perform lung impedance measurements and about sending the measurements by SMS text messages to a study mobile phone used by the HF study nurse. The HF study nurse was contacted if the participant had problems with bad electrode connections or other technical problems (for example, difficulty sending SMS data from home). Study participants and informal caregivers were instructed on how to recognize HF symptoms and signs through the use of a paper handout with textual explanations, color-coded information related to HF condition, and prominent contact information.

Symptom Diary

Memorial Symptom Assessment Scale-Heart Failure

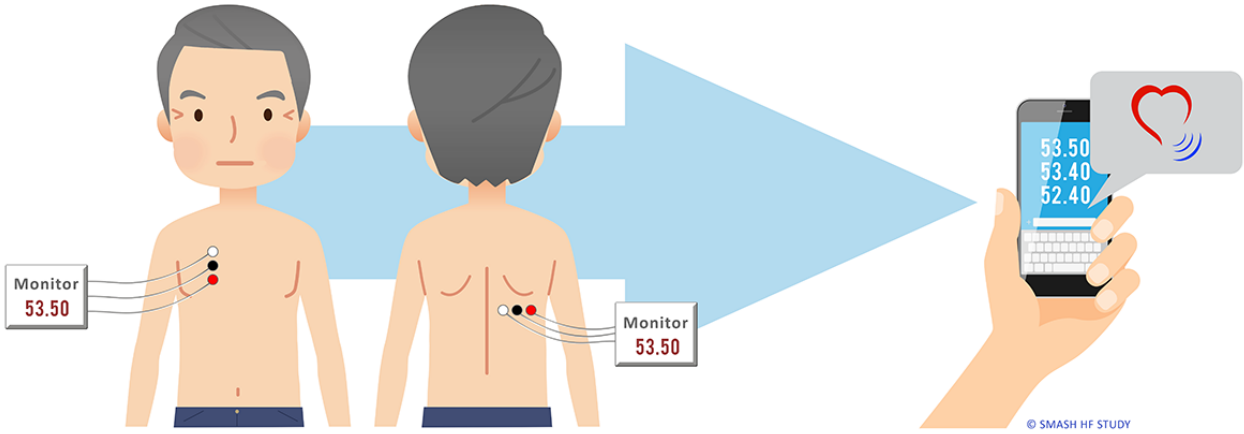
Participants rated their symptoms using an adapted version of the Memorial Symptom Assessment Scale-Heart Failure (MSAS-HF) [26], which was modified from the original Memorial Symptom Assessment Scale for cancer patients [27]. The MSAS-HF contains a list of 32 symptoms that HF patients might experience during the most recent 7 days. Participants in our study responded first about the *presence* or absence of each listed symptom and then scored the symptom *frequency* on a Likert scale from 1 to 5 (from rarely to all the time). Symptom *severity* was also scored on a Likert scale from 1 to 5 (from mild to extremely), and symptom *distress* was scored on a Likert scale from 1 to 5 (from a little bit to extremely). The total score is the sum of all the symptoms present, with a minimum score of 0 and a maximum score of 32. Symptom burden scores are determined by calculating the mean frequency, severity, and distress of each symptom, with a possible maximum score of 5 for each experienced symptom. The MSAS-HF was used with permission from the developers [28,29]. To our knowledge, no HF patients have ever used this instrument at home for 30 days after hospital discharge. Translation and cultural adoption of the original MSAS-HF from English to Norwegian and from English to Lithuanian were performed using principles of good practice for the translation and cultural adaption process for patient-reported outcomes [30]. A total of 5 nonparticipating HF patients and their spouses in Norway and Lithuania informally evaluated the translated Norwegian and Lithuanian versions of the MSAS-HF for readability and clarity. Internal consistency of the original MSAS-HF had a Cronbach alpha ranging from .73 to .91 [28,29].

Measurement of Lung Impedance

The participants and informal caregivers measured lung impedance at home daily using a CardioSet Edema Guard Monitor (Model 0001; CardioSet Medical, Ltd), a noninvasive impedance monitor developed to measure pulmonary congestion or edema in HF patients [20,31]. The patient's lung fluid status was assessed by the new impedance technique when 100 kHz alternating electrical current was passing through the patient's chest, and a unique algorithm calculated lung impedance. As HF patients may have pulmonary congestion or edema at discharge, we needed to determine the baseline or *dry* lung impedance of each participant as a reference value. This baseline value was based on each participant's age, gender, height, weight, NYHA class, and lung impedance measurements [19]. Lung impedance measurement and dry baseline lung impedance values were used to calculate the lung impedance ratio (ΔLIR=[current LI/BLI]−1×100%). Negative ΔLIR values

reflect the degree of congestion, with more negative values translating to more congestion. For ΔLIRs between 0% and −18%, change in therapy is not required; however, a larger decrease of −18% to −24% is an indication to optimize treatment. Hospitalization may be necessary if the ΔLIR decreases more than −24% [20]. In this study, if a ΔLIR at discharge was −24% or more and decreased further, the HF study nurse contacted the patients by phone asking about their condition. If necessary, the nurse told the participant to seek help according to the study protocol. Study participants performed the daily measurement in a sitting position each morning with help from their informal caregiver. The caregiver helped place the Edema Guard Monitor electrodes on the same body location every day for the 30 days. A total of 3 lung impedance measurements were taken daily in line with the developer's recommendations and sent to the HF study nurse by SMS text messages or phone calls. The HF study nurse calculated the mean impedance from the 3 lung impedance values and the ΔLIR each day for every participant. The Edema Guard Monitor consists of 6 electrodes and a device for measurements. The informal caregiver placed the 3 electrodes vertically on the right side of the chest, 4.5 cm from the midline of the sternum, with the upper electrode attached precisely under the clavicle. Next, the informal caregiver placed the 3 remaining electrodes horizontally across the lower edge of the right scapula, with the most leftward electrode located at the crossing point of the horizontal line with the spine [19,20]. Figure 1 shows a schematic illustration of the locations and positions of the electrodes. Here, an HF patient is shown performing the lung impedance measurement and sending the resulting data from home.

Figure 1. Measurement of noninvasive lung impedance at home and sending impedance data to the study center by SMS. Schematic illustration shows the correct placement of the 6 Edema Guard Monitor electrodes on an HF patient, measurement of impedance, and how daily measurement data are sent by mobile phone (SMS).



On-Demand Medication

On-demand medication was prescribed, for example, additional diuretics for some of the included participants to be used at home if necessary. Every day, these HF patients recorded their use of any of these on-demand medications in their diary.

Questionnaire

The European Heart Failure Self-care Behavior Scale (EHFScBS) is a self-rating instrument that measures HF-related self-care. It comprises 9 items that are self-scored on a 5-point Likert scale. A standardized score ranges from 0 to 100 results, with higher scores indicating better HF self-care [32-34]. The EHFScBS includes 4 items on consulting behavior and 5 items on other kinds of self-care behavior. The English version of the EHFScBS was translated into Norwegian and Lithuanian [30], as done for the MSAS-HF. Study participants completed the EHFScBS once at discharge and once 30 days later at the outpatient clinic. The EHFScBS-9 item has good internal consistency, having a Cronbach alpha coefficient between .68 and .87 [32]; in this study, the Cronbach alpha coefficient was .87.

Clinical Examination, Screening for Depression, and Assessment of Comorbidity

A standard clinical examination was performed at hospital discharge and at the outpatient clinic 30 days later (Table 1) by a cardiologist overseeing the patients care. An HF study nurse noted the patients' current medication, current NYHA class, and body weight and carefully examined the patients' jugular vein, ankles, legs, and feet for any signs of HF and fluid retention. Blood was also collected for standard laboratory tests. Participants completed the Hospital Anxiety and Depression Scale questionnaire at discharge to assess the presence of major depression, which would have been exclusionary. The language-appropriate version was used. Major depression was indicated by a score of 11 or greater [35]. Comorbidity at discharge was assessed using the Charlson Comorbidity Index (CCI). The CCI was obtained by patient interview and medical record review [36]. A total of 14 components of comorbidity are presented in the CCI, with severity ranging from value of 1, 2, 3, or 6. HF was not registered as a comorbidity for our participants.

Data Analysis

EpiData Entry (EpiData Software, 2017) was used for entering and managing data to optimize data accuracy and entry across the 2 countries. This step is in line with recommendation from the data protection managers at both university hospital sites. Raw data were converted from the EpiData Entry format to the SPSS format for analysis. Variables were either categorical or continuous and were presented as counts, percentages, means and SDs, or median and IQRs. Missing data from the symptom diary were not included in the analysis [27] and reported lung impedance data that were collected during an obvious suboptimal electrode connection or lung impedance data for which there was an intraday impedance difference of 3 Ω or more were rejected from analysis [20]. We rejected 2.3% (6/262) of lung impedance measurements for these reasons. Comparison of median self-care behavior at discharge and 30 days later was performed using Related–Samples Wilcoxon signed-rank test. A *P* value <.05 was considered statistically significant [34]. Data were analyzed using SPSS Statistics for Windows, version 25.0 (IBM Corp, released 2017).

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from Regional Committees for Medical and Health Research Ethics in Norway (REC number: 2014/1890) and Vilnius Regional Ethics Committee in Lithuania (approval number: 158200-15-766-280). All participants received verbal and written information about the purpose of the study, signed a written informed consent form before participation, and were free to withdraw from the study at any time. The daily lung impedance measurements were securely sent to the HF study nurse in each country using dedicated study mobile phones. Exchanged study data between the 2 university hospital study sites were deidentified with files encrypted in an email, which required a separate SMS code to be opened. This privacy assurance of research subject data and identity is in accordance with requirements of the data protection officer at the university hospital study sites.

The Edema Guard Monitor lung impedance measuring devices were purchased from the company CardioSet Medical Ltd, Matan, Israel, and researchers and the company signed a written contract with no obligations to the company.

Results

Participant Data

A total of 10 participants with HF in NYHA classes III and IV at inclusion (5 from each country) were recruited. The participants' mean age was 64.5 years, 8 participants had comorbidities, 1 participant was female, and most of the participants lived with their spouse or children (Table 2). None of the participants had self-reported major depression or advanced chronic renal failure. After the 30-day home assessment period, 3 HF patients were classified to be in NYHA class II (at the outpatient clinic). The participants' demographics and HF characteristics at discharge and after the 30-day period are presented in Table 2.

Table 2. Participants' demographics and clinical characteristics when discharged and 30 days later at the outpatient clinic (N=10).

Characteristics of participants	Discharge	Outpatient clinic
Age (years), mean (range)	64.5 (37-85)	—[a]
Gender, n		
Male	9	—
Education, n		
Less than high school	1	—
Informal caregiver at home (help with lung impedance measurement), n		
Spouse	7	—
Grown children	2	—
Nurse	1	—
Work status, n		
Full time	3	—
Retired, disability pension	7	—
HF[b] diagnosis, n		
Ischemic HF	7	—
Dilated	3	—
HF >1 year	7	—
HF <1 year	3	—
New York Heart Associations classes, n		
II	0	3
III	8	6
IV	2	1
Implantable cardioverter-defibrillator, n	3	3
Cardiac resynchronization therapy with pacemaker or defibrillation, n	3	3
Jugular venous pressure, n	1	1
Ankle swelling, n	5	2
Systolic BP[c] (mm Hg), mean (SD)	113.3 (17.5)	119.2 (19.7)
Diastolic BP (mm Hg), mean (SD)	74.7 (13.8)	77.9 (12.4)
Heart rate, mean (SD)	76.5 (13.6)	75.3 (12.0)
HF medication, n		
Angiotensin-converting enzyme inhibitors or angiotensin receptor blockers	9	10
Beta blockers	9	9
Loop diuretics	10	10
Thiazides diuretics	1	1
Mineralocorticoid antagonist	7	5
Ivabradine	1	2
Nitrates	1	1
Charlson Comorbidity Index, n		
No comorbidity	2	—
Low (1-2)	3	—
Medium (3-4)	5	—

Characteristics of participants	Discharge	Outpatient clinic
Blood values, mean (SD)		
Hemoglobin, g/dL (reference: 13.4-17.0)	14.7 (1.7)	14.6 (1.1)
Creatinine, mmol/L (reference: 60-105)	104.3 (23.7)	122.2 (59.4)
N-terminal pro b-type natriuretic peptide, pg/mL	2200 (1408)	2910 (2788)
Estimate of glomerular filtration rate, mL/min/1.73m^2	66.3 (19.3)	—

[a]Number same at discharge and 30 days later at the outpatient clinic.

[b]HF: heart failure.

[c]BP: blood pressure.

Daily Use of a Symptom Diary and a Lung Impedance Measurement Device at Home

Participants used the symptom diary daily, and with support from their caregivers, they measured lung impedance daily. A total of 7 participants were able to provide data for the full 30-day assessment period. For impedance data, 262 of 300 (87%) recorded measurements were successfully made and sent. For the symptom diary, 1332 of 1500 (89%) entries were successfully made. Missing data were mainly from 2 participants who were classified as NYHA class IV. These participants provided diary and lung impedance data for 19 of 30 days (57%) and 22 of 30 days (63%), respectively. These 2 participants had a few missing lung impedance measurements because of issues with poor signal quality presumably related to suboptimal electrode connections, despite performing more than 3 required measurements and receiving advice from the HF study nurse. A third participant did not have home caregiver support for measuring lung impedance after day 26; thus, participation was terminated on day 26 for that participant. Moreover, 2 participants lived alone, but 1 of them moved in with a family member during the 30-day home assessment period. The other participant received support from an HF study nurse every day to apply the electrodes, although the participant performed the measurement.

Heart Failure Symptoms During 30-Day Postdischarge Assessment

Prevalence and burden assessment of the 5 symptoms selected from the MSAS-HF varied among the participants (Table 3). Out of the 10 participants, 7 experienced shortness of breath for at least one of the 30 days, and of these participants, 4 experienced it every day of the 30-day period. Out of the 10 participants, 8 reported feeling lack of energy for at least 1 day, and 3 reported lack of energy every day. The least commonly occurring symptom was swelling of hands, legs, or feet. Out of the 10 participants, 3 (30%) reported swelling of the extremities on at least one of the 30 days; no one reported experiencing swelling every day.

The total burden score (mean of the frequency, severity, and distress scores) of *shortness of breath* and *lacking energy* was higher than the total burden score of *difficulty breathing lying flat*; *wake up breathless*; or *swelling of hands, legs, or feet*. Moreover, 1 participant did not experience any of the 5 HF symptoms or signs during the 30-day period. Details of HF symptom prevalence and total burden scores are presented in Table 3.

A total of 9 participants experienced symptoms indicating fluid accumulation for 1 or more days. They reported feeling *shortness of breath* for 5 to 26 days and *lack of energy* for 1 to 28 days.

Table 3. The number of participants experiencing (prevalence) and total burden score (mean of the frequency, severity, and distress scores) of the 5 heart failure symptoms and signs during the 30-day postdischarge assessment period (N=10).

Symptom	Prevalence			Total burden score					
	None of the days, n	1 day or more, n	Every day, n	1 day or more			Every day		
				Mean (SD)	Minimum	Maximum	Mean (SD)	Minimum	Maximum
Shortness of breath	3	7	4	2.9 (0.8)	2.3	3.8	2.8 (0.5)	2.3	3.4
Difficulty breathing lying flat	4	6	1	2.4 (1.3)	1.2	4.6	2.8(-[a])	2.8	2.8
Wake up breathless	5	5	1	1.6 (0.6)	1.0	2.4	2.9 (-)	2.9	2.9
Lack of energy	2	8	3	2.2 (1.3)	1.3	4.4	3.6 (1.1)	2.6	4.8
Swelling of extremities	7	3	0	1.7 (0.6)	1.1	2.3	0 (-)	0	0

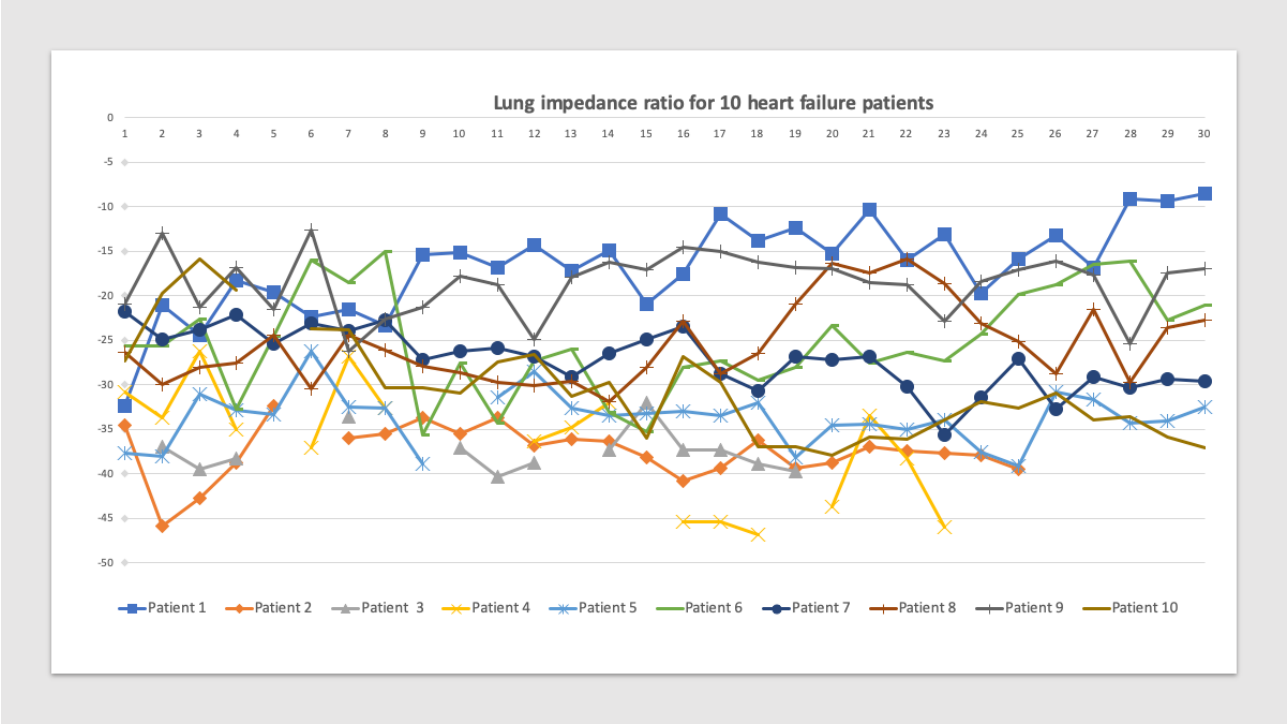
[a]Only 1 participant had the symptom every day.

Noninvasive Lung Impedance Ratio During the 30 Days After Hospitalization

At hospital discharge, individual ΔLIRs ranged from −20.9% to −37.7%. Thus, on day 1 of the study, all participants may have presence of pulmonary edema. The participants' LIRs over time are presented in Figure 2.

The number of days the patients participated in the self-assessment varied because of technical or health-related reasons. Patient number 2 had 26 days of data, patient number 3 had 19 days of data, and patient number 4 had 23 days of data (Figure 2). Patient number 4 and 10 were temporarily hospitalized during the 30 days because of a worsening HF condition (patient number 4) or acute renal failure (patient number 10). Before readmission, the ΔLIR for patient number 4 was −46.81%, and the ΔLIR for patient number 10 was −19.4%. At the end of the 30-day home assessment period, individual lung impedance values ranged from −8.5% to −39.6%. These values indicate that some participants had minimal pulmonary fluid present (ΔLIRs between 0% and −18%), whereas others had actionable edema requiring adjustment of medication and possible hospitalization. Moreover, 4 patients required medication adjustment during the 30-day study period. Patient number 3 was instructed to self-administer diuretics at home. Patient number 7, 9, and 10 were invited to an outpatient consultation (Lithuanian study site) with a cardiologist for medication adjustment.

Figure 2. Individual lung impedance ratios (ΔLIRs) during the 30-day postdischarge at-home assessment period. Each color or symbol represents the ΔLIR time series for 1 of the 10 participants. More negative ΔLIRs translate to more pulmonary congestion. Missing data (interrupted time series) resulted from a poor electrode connection, an intervening hospitalization, or a participant temporarily lacking help from their home caregiver to perform measurements. An Edema Guard Monitor (CardioSet Medical) lung impedance device was used to measure lung impedance (see the Methods section for calculation of ΔLIR).



Self-Care Behavior

Self-care behavior improved from the time they were discharged to the end of the 30-day assessment period. The total score on the EHFScBS at the end was significantly better (*P*=.007) than that at hospital discharge. The subscale score on consulting behavior also significantly improved (*P*=.049) after 30 days. At the item level, the most frequently performed self-care behavior (score 5) was *taking medicationas prescribed*, and the least frequently performed self-care behavior (score 1) was *to exercise regularly*. Details of the self-care behavior results are presented in Table 4.

Table 4. Comparison of heart failure participants' self-care behavior at hospital discharge and 30 days later at the outpatient clinic as assessed with the European Heart Failure Self-care Behavior Scale (European Heart Failure Self-care Behavior Scale-9 comprises 9 items scored on a Likert scale of 1-5, with a higher score indicating better adherence to the given behavior; N=10).

Behavior	Discharge, median (IQR)	30 days, median (IQR)	*P* value
I weigh myself every day	4.5 (1.8-5)	5 (4.8-5)	.04
If my shortness of breath increases, I contact my doctor/nurse	2.5 (1-5)	5 (3-5)	.07
If my feet/legs become swollen, I contact my doctor/nurse	3.5 (1-5)	5 (3.8-5)	.08
If I gain 2 kg in 1 week, I contact my doctor/nurse	2.5 (1- 4.3)	4.5 (2.8-5)	.14
I limit the amount of fluids I drink (<1.5 to 2 L per day)	3.5 (2-5)	4.5 (4-5)	.06
If I experience increased fatigue, I contact my doctor/nurse	2.5 (1.8-3.3)	5 (1.8-5)	.07
I eat a low-salt diet	3.5 (2-5)	5 (4-5)	.03
I take my medication as prescribed	5 (4.8-5)	5 (5-5)	.18
I exercise regularly	2 (1-3)	3.5 (2-5)	.01
European Heart Failure Self-care Behavior Scale, 9 items, total standardized score[a]	56 (22-75)	81 (72-98)	.007
Consulting behavior	55 (38-79)	88 (58-100)	.049

[a]Standardized or summed scores on the European Heart Failure Self-care Behavior Scale, 9 items, ranged from 0 to 100, with higher scores indicating better self-care behavior.

Discussion

Principal Findings

Our main finding was that HF patients were able to successfully use a symptom diary and measure aided noninvasive lung impedance at home to self-monitor and report HF symptoms during the 30-day postassessment. This is the first study in which HF patients would use the Edema Guard Monitor to measure their lung impedance on a daily basis for 30 days following hospital discharge to home. Previously, the Edema Guard Monitor had been used by health care professionals once a month at an outpatient clinic with HF patients [20]. Lung impedance measurements were aided by informal caregivers. Study participants significantly improved their self-care behavior during this period. None of the 10 participants or their caregivers had previous experience with the symptom diary or with the noninvasive lung impedance instrument; yet with minimal training, they successfully completed and sent 87% of impedance measurements. Finally, only 2 participants (patient number 4 and 10) had to be readmitted during the 30-day assessment period. Both readmissions were triggered by self-detected worsening conditions (ΔLIR for patient number 4 was −46.81%, and ΔLIR for patient number 10 was −19.4%).

The age range of our participants was 37 to 85 years, and our findings are in line with some reports showing that older patients accept and manage well when using new technology [37]. However, other studies showed that familiarity is key in older patients' acceptance of new technology at home [38]. All participants in our study had daily contact with the HF study nurse by SMS text messages or phone calls. One reason why our participants managed well may be that they had remote access (phone) to the HF study nurse if they had a poor electrode connection or any other technical problems using the lung impedance device. Another possible explanation is that the participants were actively involved in performing, recording, and sending lung impedance measurements to the HF study nurse. The HF study nurse assessed changes in measurements that participants sent, allowing them to give participants feedback for possible technical adjustments in the device. Participants in other studies likely did not have daily access to professional feedback [39,40]. This kind of access to professional help is not part of routine HF follow-up in Norway and Lithuania. The nurse ratio is different in the 2 countries, with 17.7 per 1000 inhabitants in Norway and 7. 7 per 1000 inhabitants in Lithuania [41].

Typically, after discharge from hospital, HF patients are referred to their general practitioner in the community health care services for follow-up and medication adjustments. In Norway, follow-up at an HF outpatient clinic requires a referral from the hospital ward or general practitioner, currently occurring only 21% of the time [42]. In Lithuania, there is no special follow-up protocol for HF patients. In this study, 3 participants received additional follow-up at the outpatient clinic because they experienced an increase in HF symptoms after medication titration and an improved ΔLIR. Additional follow-up reportedly has a positive impact on the quality of life and self-care of HF patients [43]. Most of our participants also received help from their informal caregivers to measure lung impedance during the 30-day study period. The caregivers expressed positive feelings of being actively involved in the daily measurement of lung impedance and the need for their presence every day [44]. Although some studies report data on caregiver burden from monitoring HF patients [45,46], our study sheds additional light on the caregiver relationship and even reflects a desire to be involved in the care.

Another finding was that participants' self-reported ratings on the 5 key HF symptoms and on the lung impedance measurements varied greatly. The variability in HF symptoms over the course of the assessment period stresses that constant attention is needed to achieve symptom relief. Some authors also warn that HF patients acclimate to symptoms and adapt their activities and lifestyle to reduce the HF symptoms; thus,

they might not recognize these symptoms as warning signs [47-49]. Taken together, these findings indicate that active monitoring is essential for HF patients to build awareness about HF symptoms and signs when confronted with symptoms [49]. Using a symptom diary has the potential to support self-care so that HF patients can play an active role in monitoring their HF symptoms of fluid overload after hospital discharge to home.

HF decompensation is the worsening of HF symptoms and signs and is characterized by a gradual decrease in lung impedance measurements in patients at high risk for volume overload [40]. In this study, participant number 4 exhibited gradual daily changes in lung impedance (Figure 2) and was readmitted to hospital 18 days postdischarge for HF decompensation. Another participant (number 10; Figure 2) similarly experienced a gradual change in lung impedance but was readmitted shortly after discharge for acute renal failure. In our study, gradual changes in HF symptoms appeared relatively quickly following discharge compared with those reported in other studies, which took several weeks to manifest [40]. This may be because, in our study, these patients received adjustments in their medications or because data were missing. Another reason for discrepancy of results across the HF literature relates to use of inconsistent terminology to describe pulmonary edema symptoms and to differing methods used to measure edema (eg, thoracic, intrathoracic, or bioimpedance measurements and noninvasive methods) [39,50-52].

We observed that self-care behavior of the study participants significantly improved after using the daily symptom diary and taking daily lung impedance measurements. At discharge, our participants had inadequate self-care. According to the EHFScBS, a score of 70 or higher reflects adequate self-care behavior [53], which is defined as actions to maintain life, healthy functioning, and well-being [32]. In patients with HF, inadequate self-care is an independent risk factor for adverse clinical outcomes [54]. After the 30-day home assessment period, our participants' self-care behavior significantly improved, as their EHFScBS scores reached or exceeded the 70-point threshold, suggesting adequate self-care. The professional activities of the HF study nurse could be one reason for this improvement in self-care among our participants. The nurse was responsible for following up each participant daily. This kind of follow-up is not part of usual follow-up for HF patients in either country. Other studies have reported that self-care is troublesome for HF patients after hospital discharge to home (eg, coordinating own care and medications and requiring information and training) [55]. Moreover, HF patients who use invasive lung impedance measuring devices and who report better self-care also have fewer episodes of decrease in lung impedance [56]. Another potential reason for the improved self-care behavior we observed is that our study enabled our participants to establish a new daily routine of filling in a symptom diary and measuring lung impedance. Routines and habits are factors known to positively influence self-care [6]. In addition, the education and the handout summarizing HF warning signs and the emergency contact that the participants received could have contributed to participants' improvement in consulting behavior. Using tools to support HF patients can, over time, change their self-care behavior for the better. Indeed, in HF patients, regular assessments are encouraged [57]. This study has demonstrated that using new self-care tools, an HF symptom diary, and a noninvasive impedance device, during the first 30 days after hospital discharge can contribute positively to better self-care.

Strengths and Limitations

One limitation is that only 10 patients participated in this study. Nonetheless, we were able to obtain a substantial amount of data from these participants, and these data provided us with an in-depth understanding of the trajectory of HF symptoms during the first month at home after hospitalization. The variability of the data for HF symptoms and LIR, in addition to our small sample, made it not advisable to perform an analysis for possible associations among these variables. Another limitation is the use of an unvalidated version of the MSAS-HF for the purpose of daily self-monitoring for 30 days at home. However, the unvalidated version did include 5 HF symptoms of fluid overload that are also recommended in guidelines [1]. Moreover, before our study, the MSAS-HF had never been used every day for 30 days by HF patients after hospital discharge. Thus, these data represent pilot normative data for the daily use of MSAS-HF.

Conclusions

This is the first study showing that discharged HF patients can successfully use a symptom diary combined with measurement of their noninvasive lung impedance at home to self-monitor on a daily basis, with support from informal caregivers. We found a significant improvement in the participants' self-care behavior in the period between hospital discharge and end of the 30-day home assessment. These self-care tools can be used by nurses in clinical practice to educate HF patients and their informal caregivers before discharge to home, which could also reduce readmissions. For future research, a larger study population is needed to determine whether self-monitoring of particular HF symptoms at home and LIR are predictive of HF deterioration. Early detection of worsening HF at home will aid health care professionals in providing better HF care.

Acknowledgments

This work was supported by the Nordic funding organization, NordForsk. The study was funded under Nordic Programme on Health and Welfare, project number: 76015, "Symptom monitoring after hospitalization in patients with advanced heart failure–a Nordic-Baltic study." The authors wish to thank all participants, their informal caregivers, and health care professionals who participated in the study. The authors thank the photographer/graphic designer Per Marius Didriksen at Oslo University Hospital for designing and producing the layout of study material and for Figure 1 in this publication.

Conflicts of Interest

None declared.

References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Authors/Task Force Members, Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016 Aug;18(8):891-975 [FREE Full text] [doi: 10.1002/ejhf.592] [Medline: 27207191]
2. Jaarsma T, Strömberg A, Gal TB, Cameron J, Driscoll A, Duengen H, et al. Comparison of self-care behaviors of heart failure patients in 15 countries worldwide. Patient Educ Couns 2013 Jul;92(1):114-120. [doi: 10.1016/j.pec.2013.02.017] [Medline: 23579040]
3. Lee KS, Moser DK, Dracup K. Relationship between self-care and comprehensive understanding of heart failure and its signs and symptoms. Eur J Cardiovasc Nurs 2018 Aug;17(6):496-504. [doi: 10.1177/1474515117745056] [Medline: 29192794]
4. Riegel B, Jaarsma T, Strömberg A. A middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2012;35(3):194-204. [doi: 10.1097/ANS.0b013e318261b1ba] [Medline: 22739426]
5. Lee CS, Moser DK, Lennie TA, Riegel B. Event-free survival in adults with heart failure who engage in self-care management. Heart Lung 2011;40(1):12-20 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.12.003] [Medline: 20561885]
6. Jaarsma T, Cameron J, Riegel B, Stromberg A. Factors related to self-care in heart failure patients according to the middle-range theory of self-care of chronic illness: a literature update. Curr Heart Fail Rep 2017 Apr;14(2):71-77 [FREE Full text] [doi: 10.1007/s11897-017-0324-1] [Medline: 28213768]
7. Spaling MA, Currie K, Strachan PH, Harkness K, Clark AM. Improving support for heart failure patients: a systematic review to understand patients' perspectives on self-care. J Adv Nurs 2015 Nov;71(11):2478-2489. [doi: 10.1111/jan.12712] [Medline: 26084885]
8. Clark AM, Savard LA, Spaling MA, Heath S, Duncan AS, Spiers JA. Understanding help-seeking decisions in people with heart failure: a qualitative systematic review. Int J Nurs Stud 2012 Dec;49(12):1582-1597. [doi: 10.1016/j.ijnurstu.2012.05.010] [Medline: 22721677]
9. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. ESC Heart Fail 2014 Sep;1(1):4-25 [FREE Full text] [doi: 10.1002/ehf2.12005] [Medline: 28834669]
10. Dharmarajan K, Chaudhry SI. New approaches to reduce readmissions in patients with heart failure. JAMA Intern Med 2016 Mar;176(3):318-320. [doi: 10.1001/jamainternmed.2015.7993] [Medline: 26857025]
11. Khan H, Greene SJ, Fonarow GC, Kalogeropoulos AP, Ambrosy AP, Maggioni AP, EVEREST Trial Investigators. Length of hospital stay and 30-day readmission following heart failure hospitalization: insights from the EVEREST trial. Eur J Heart Fail 2015 Oct;17(10):1022-1031 [FREE Full text] [doi: 10.1002/ejhf.282] [Medline: 25960401]
12. Rocha BM, Falcão LM. Acute decompensated heart failure (ADHF): A comprehensive contemporary review on preventing early readmissions and postdischarge death. Int J Cardiol 2016 Nov 15;223:1035-1044. [doi: 10.1016/j.ijcard.2016.07.259] [Medline: 27592046]
13. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. J Am Med Assoc 2013 Jan 23;309(4):355-363 [FREE Full text] [doi: 10.1001/jama.2012.216476] [Medline: 23340637]
14. Canepa M, Straburzynska-Migaj E, Drozdz J, Fernandez-Vivancos C, Pinilla JM, Nyolczas N, ESC-HFA Heart Failure Long-Term Registry Investigators. Characteristics, treatments and 1-year prognosis of hospitalized and ambulatory heart failure patients with chronic obstructive pulmonary disease in the European Society of Cardiology Heart Failure Long-Term Registry. Eur J Heart Fail 2018 Jan;20(1):100-110 [FREE Full text] [doi: 10.1002/ejhf.964] [Medline: 28949063]
15. Köberich S. Fostering self-care behaviours through symptom diary use? An exploratory, cross-sectional study about the use of and attitude towards a symptom diary of patients with heart failure. J Res Nurs 2016;21(7):530-541. [doi: 10.1177/1744987116674256]
16. Park LG, Dracup K, Whooley MA, McCulloch C, Jin C, Moser DK, et al. Symptom diary use and improved survival for patients with heart failure. Circ Heart Fail 2017 Nov;10(11):pii: e003874 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.117.003874] [Medline: 29158435]
17. Lee S, Riegel B. State of the science in heart failure symptom perception research: an integrative review. J Cardiovasc Nurs 2018;33(3):204-210. [doi: 10.1097/JCN.0000000000000445] [Medline: 28858886]
18. Jones CD, Holmes GM, DeWalt DA, Erman B, Wu J, Cene CW, et al. Self-reported recall and daily diary-recorded measures of weight monitoring adherence: associations with heart failure-related hospitalization. BMC Cardiovasc Disord 2014 Jan 31;14:12 [FREE Full text] [doi: 10.1186/1471-2261-14-12] [Medline: 24483776]

19. Shochat M, Shotan A, Blondheim DS, Kazatsker M, Dahan I, Asif A, et al. Derivation of baseline lung impedance in chronic heart failure patients: use for monitoring pulmonary congestion and predicting admissions for decompensation. J Clin Monit Comput 2015 Jun;29(3):341-349. [doi: 10.1007/s10877-014-9610-6] [Medline: 25193676]
20. Shochat MK, Shotan A, Blondheim DS, Kazatsker M, Dahan I, Asif A, et al. Non-invasive lung IMPEDANCE-guided preemptive treatment in chronic heart failure patients: a randomized controlled trial (IMPEDANCE-HF trial). J Card Fail 2016 Sep;22(9):713-722. [doi: 10.1016/j.cardfail.2016.03.015] [Medline: 27058408]
21. World Bank Open Data. 2018. GDP per capita (current US$) - Lithuania, Ukraine, Latvia, Estonia, Norway URL: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=LT-UA-LV-EE-NO [accessed 2019-06-01]
22. Murauskiene L, Janoniene R, Veniute M, van Ginneken E, Karanikolos M. Lithuania: health system review. Health Syst Transit 2013;15(2):1-150 [FREE Full text] [Medline: 23902994]
23. Ringard A, Sagan A, Sperre Saunes I, Lindahl AK. Norway: health system review. Health Syst Transit 2013;15(8):1-162 [FREE Full text] [Medline: 24434287]
24. Higienos Institutas. 2018. Prevalence - number of ill persons URL: https://stat.hi.lt/default.aspx?report_id=256 [accessed 2019-08-22]
25. Halvorsen S, Odegaard KM, Lirhus S, Arneberg F, Melberg HO. Incidence, prevalence and survival in heart failure: a nationwide registry study from 2011-2016. Eur Heart J 2018;39(Suppl 1):1388. [doi: 10.1093/eurheartj/ehy566.P6520]
26. Zambroski C, Lennie T, Chung M, Heo S, Smoot T, Ziegler C. Use of the Memorial Symptom Assessment Scale-Heart Failure in heart failure patients. Circulation 2004;110(Supplement 111 (17)):739 [FREE Full text]
27. Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 1994;30A(9):1326-1336. [doi: 10.1016/0959-8049(94)90182-1] [Medline: 7999421]
28. Lee KS, Moser DK. Heart failure symptom measures: critical review. Eur J Cardiovasc Nurs 2013 Oct;12(5):418-428. [doi: 10.1177/1474515112473235] [Medline: 23303769]
29. Zambroski CH, Moser DK, Bhat G, Ziegler C. Impact of symptom prevalence and symptom burden on quality of life in patients with heart failure. Eur J Cardiovasc Nurs 2005 Sep;4(3):198-206. [doi: 10.1016/j.ejcnurse.2005.03.010] [Medline: 15916924]
30. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (pro) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
31. Shochat M, Charach G, Meyler S, Meisel S, Weintraub M, Mengeritsky G, et al. Prediction of cardiogenic pulmonary edema onset by monitoring right lung impedance. Intensive Care Med 2006 Aug;32(8):1214-1221. [doi: 10.1007/s00134-006-0237-z] [Medline: 16775717]
32. Jaarsma T, Arestedt KF, Mårtensson J, Dracup K, Strömberg A. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument. Eur J Heart Fail 2009 Jan;11(1):99-105 [FREE Full text] [doi: 10.1093/eurjhf/hfn007] [Medline: 19147463]
33. Sedlar N, Socan G, Farkas J, Mårtensson J, Strömberg A, Jaarsma T, et al. Measuring self-care in patients with heart failure: a review of the psychometric properties of the European Heart Failure Self-Care Behaviour Scale (EHFScBS). Patient Educ Couns 2017 Jul;100(7):1304-1313. [doi: 10.1016/j.pec.2017.02.005] [Medline: 28209470]
34. Vellone E, Jaarsma T, Strömberg A, Fida R, Årestedt K, Rocco G, et al. The European Heart Failure Self-care Behaviour Scale: new insights into factorial structure, reliability, precision and scoring procedure. Patient Educ Couns 2014 Jan;94(1):97-102. [doi: 10.1016/j.pec.2013.09.014] [Medline: 24148605]
35. Snaith RP. The hospital anxiety and depression scale. Health Qual Life Outcomes 2003 Aug 1;1:29 [FREE Full text] [doi: 10.1186/1477-7525-1-29] [Medline: 12914662]
36. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
37. Lind L, Carlgren G, Karlsson D. Old-and with severe heart failure: telemonitoring by using digital pen technology in specialized homecare: system description, implementation, and early results. Comput Inform Nurs 2016 Aug;34(8):360-368. [doi: 10.1097/CIN.0000000000000252] [Medline: 27223309]
38. Holender A, Sutton S, De Simoni A. Opinions on the use of technology to improve tablet taking in >65-year-old patients on cardiovascular medications. J Int Med Res 2018 Jul;46(7):2754-2768 [FREE Full text] [doi: 10.1177/0300060518770578] [Medline: 29730949]
39. Riegel B, Dickson VV, Lee CS, Daus M, Hill J, Irani E, et al. A mixed methods study of symptom perception in patients with chronic heart failure. Heart Lung 2018;47(2):107-114 [FREE Full text] [doi: 10.1016/j.hrtlng.2017.11.002] [Medline: 29304990]
40. Gudmundsson K, Lyngå P, Rosenqvist M, Braunschweig F. Monitoring of daily body weight and intrathoracic impedance in heart failure patients with a high risk of volume overload decompensation. Clin Cardiol 2016 Aug;39(8):446-452 [FREE Full text] [doi: 10.1002/clc.22547] [Medline: 27175605]

41. The Organisation for Economic Co-operation and Development Data. 2018. Nurses (indicator) URL: https://data.oecd.org/healthres/nurses.htm [accessed 2018-11-04] [WebCite Cache ID 73fo11GpD]
42. The National Norwegian Heart Failure Registry (NNHFR). National service environment for quality medical records. 2017. Årsrapport 2017 med plan for forbedringstiltak [Annual report] URL: https://www.kvalitetsregistre.no/sites/default/files/6_arsrapport_2017_norsk_hjertesviktregister_0.pdf [accessed 2019-06-01]
43. Sezgin D, Mert H, Özpelit E, Akdeniz B. The effect on patient outcomes of a nursing care and follow-up program for patients with heart failure: a randomized controlled trial. Int J Nurs Stud 2017 May;70:17-26. [doi: 10.1016/j.ijnurstu.2017.02.013] [Medline: 28214615]
44. Lycholip E, Aamodt IT, Celutkiene J, Lie I, Jaarsma T, Stromberg A. European Society of Cardiology 365. 2019. Experiences of Heart Failure Patients and Their Informal Caregivers With Using Monitoring of Lung Impedance at Home URL: https://esc365.escardio.org/Congress/EuroHeartCare-2019/Poster-Session-2-Cardiovascular-Nursing/192461-experiences-of-heart-failure-patients-and-their-informal-carers-with-using-monitoring-of-lung-impedance-at-home#abstract [accessed 2019-06-01]
45. Strøm A, Andersen KL, Korneliussen K, Fagermoen MS. Being 'on the alert' and 'a forced volunteer': a qualitative study of the invisible care provided by the next of kin of patients with chronic heart failure. J Multidiscip Healthc 2015;8:271-277 [FREE Full text] [doi: 10.2147/JMDH.S82239] [Medline: 26082643]
46. Luttik ML, Jaarsma T, Strömberg A. Changing needs of heart failure patients and their families during the illness trajectory: a challenge for health care. Eur J Cardiovasc Nurs 2016 Aug;15(5):298-300. [doi: 10.1177/1474515116653536] [Medline: 27255508]
47. Alpert CM, Smith MA, Hummel SL, Hummel EK. Symptom burden in heart failure: assessment, impact on outcomes, and management. Heart Fail Rev 2017 Jan;22(1):25-39 [FREE Full text] [doi: 10.1007/s10741-016-9581-4] [Medline: 27592330]
48. Ängerud KH, Boman K, Brännström M. Areas for quality improvements in heart failure care: quality of care from the family members' perspective. Scand J Caring Sci 2018 Mar;32(1):346-353. [doi: 10.1111/scs.12468] [Medline: 28543624]
49. Riegel B, Jaarsma T, Lee CS, Strömberg A. Integrating symptoms into the middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2019;42(3):206-215 [FREE Full text] [doi: 10.1097/ANS.0000000000000237] [Medline: 30475237]
50. Domenichini G, Rahneva T, Diab IG, Dhillon OS, Campbell NG, Finlay MC, et al. The lung impedance monitoring in treatment of chronic heart failure (the LIMIT-CHF study). Europace 2016 Mar;18(3):428-435. [doi: 10.1093/europace/euv293] [Medline: 26683599]
51. Smeets CJ, Vranken J, van der Auwera J, Verbrugge FH, Mullens W, Dupont M, et al. Bioimpedance alerts from cardiovascular implantable electronic devices: observational study of diagnostic relevance and clinical outcomes. J Med Internet Res 2017 Nov 23;19(11):e393 [FREE Full text] [doi: 10.2196/jmir.8066] [Medline: 29170147]
52. Hawkins NM, Virani SA, Sperrin M, Buchan IE, McMurray JJV, Krahn AD. Predicting heart failure decompensation using cardiac implantable electronic devices: a review of practices and challenges. Eur J Heart Fail 2016 Aug;18(8):977-986 [FREE Full text] [doi: 10.1002/ejhf.458] [Medline: 26663507]
53. Wagenaar KP, Broekhuizen BD, Rutten FH, Strömberg A, van Stel HF, Hoes AW, et al. Interpretability of the European Heart Failure Self-care Behaviour scale. Patient Prefer Adherence 2017;11:1841-1849 [FREE Full text] [doi: 10.2147/PPA.S144915] [Medline: 29138538]
54. Kato N, Kinugawa K, Nakayama E, Tsuji T, Kumagai Y, Imamura T, et al. Insufficient self-care is an independent risk factor for adverse clinical outcomes in Japanese patients with heart failure. Int Heart J 2013;54(6):382-389 [FREE Full text] [doi: 10.1536/ihj.54.382] [Medline: 24309448]
55. Nordfonn OK, Morken IM, Bru LE, Husebø AM. Patients' experience with heart failure treatment and self-care-a qualitative study exploring the burden of treatment. J Clin Nurs 2019 May;28(9-10):1782-1793. [doi: 10.1111/jocn.14799] [Medline: 30667120]
56. Rathman LD, Lee CS, Sarkar S, Small RS. A critical link between heart failure self-care and intrathoracic impedance. J Cardiovasc Nurs 2011;26(4):E20-E26. [doi: 10.1097/JCN.0b013e3181ee28c8] [Medline: 21076309]
57. Lycholip E, Aamodt IT, Lie I, Šimbelytė T, Puronaitė R, Hillege H, et al. The dynamics of self-care in the course of heart failure management: data from the IN TOUCH study. Patient Prefer Adherence 2018;12:1113-1122 [FREE Full text] [doi: 10.2147/PPA.S162219] [Medline: 29983549]

Abbreviations

CCI: Charlson Comorbidity Index
EHFScBS: European Heart Failure Self-care Behavior Scale
HF: heart failure
LIR: lung impedance ratio
MSAS-HF: Memorial Symptom Assessment Scale-Heart Failure
NYHA: New York Heart Association

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by H Mehdizadeh, T Gal, M Smith; comments to author 17.08.19; revised version received 22.08.19; accepted 23.09.19; published 07.01.20.

Please cite as:
Aamodt IT, Lycholip E, Celutkiene J, von Lueder T, Atar D, Falk RS, Hellesø R, Jaarsma T, Strömberg A, Lie I
Self-Care Monitoring of Heart Failure Symptoms and Lung Impedance at Home Following Hospital Discharge: Longitudinal Study
J Med Internet Res 2020;22(1):e15445
URL: https://www.jmir.org/2020/1/e15445
doi:10.2196/15445
PMID:31909717

Original Paper

The Effects of a Digital Well-Being Intervention on Patients With Chronic Conditions: Observational Study

Acacia C Parks[1], PhD; Allison L Williams[1], PhD; Gina M Kackloudis[1], MPH; Julia L Stafford[1], BA; Eliane M Boucher[1], PhD; Ryan D Honomichl[2], PhD

[1]Happify, New York, NY, United States

[2]Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Acacia C Parks, PhD
Happify
51 East 12th Street, 5th Floor
New York, NY, 10003
United States
Phone: 1 2678798387
Email: acacia@happify.com

Abstract

Background: Chronic conditions account for 75% of health care costs, and the impact of chronic illness is expected to grow over time. Although subjective well-being predicts better health outcomes, people with chronic conditions tend to report lower well-being. Improving well-being might mitigate costs associated with chronic illness; however, existing interventions can be difficult to access and draw from a single theoretical approach. Happify, a digital well-being intervention program drawing from multiple theoretical traditions to target well-being, has already been established as an efficacious means of improving well-being in both distressed and nondistressed users.

Objective: This study aimed to compare change in well-being over time after using Happify for users with and without a chronic condition.

Methods: Data were obtained from Happify users, a publicly available digital well-being program accessible via website or mobile phone app. Users work on tracks addressing a specific issue (eg, conquering negative thoughts) composed of games and activities based on positive psychology, cognitive behavioral therapy, and mindfulness principles. The sample included 821 users receiving at least 6 weeks' exposure to Happify (ranging from 42 to 179 days) who met other inclusion criteria. As part of a baseline questionnaire, respondents reported demographic information (age and gender) and whether they had any of the prespecified chronic conditions: arthritis, diabetes, insomnia, multiple sclerosis, chronic pain, psoriasis, eczema, or some other condition (450 reported a chronic condition, whereas 371 did not). Subjective well-being was assessed with the Happify Scale, a 9-item measure of positive emotionality and life satisfaction. To evaluate changes in well-being over time, a mixed effects linear regression model was fit for subjective well-being, controlling for demographics and platform usage.

Results: At baseline, users with a chronic condition had significantly lower subjective well-being (mean 38.34, SD 17.40) than users without a chronic condition (mean 43.65, SD 19.13). However, change trajectories for users with or without a chronic condition were not significantly different; both groups experienced equivalent improvements in well-being. We also found an effect for time from baseline (b=0.071; SE=0.010; P<.01) and number of activities completed (b=0.03; SE=0.009; P<.01), and a 2-way interaction between number of activities completed and time from baseline (b=0.0002; SE=0.00006; P<.01), such that completing more activities and doing so over increasingly longer periods produced improved well-being scores.

Conclusions: Data from this study support the conclusion that users with a chronic condition experienced significant improvement over time. Despite reporting lower subjective well-being on the whole, their change trajectory while using Happify was equivalent to those without a chronic condition. Consistent with past research, users who completed more activities over a longer period showed the most improvement. In short, the presence of a chronic condition did not prevent users from showing improved well-being when using Happify.

(J Med Internet Res 2020;22(1):e16211) doi:10.2196/16211

KEYWORDS

chronic illness; happiness; subjective well-being; psychology, positive; internet-based intervention; mobile apps

Introduction

Background

According to the Center for Disease Control, chronic conditions are the leading cause of death and disability in the United States. Chronic conditions affect over 40% of the US population [1], and almost 20% of individuals with chronic illnesses report activity limitations that impede their ability to complete daily tasks, including work [2]. Indeed, having a self-reported chronic condition is associated with decreased work function [3] and appears to be the strongest predictor of absenteeism and work impairment [4].

People with chronic conditions also account for the greatest use of health care services [2,5], accounting for 75% or more of health care costs [1]. Relative to people without chronic conditions, health care spending is 3 times higher for individuals with 1 chronic condition [2] and increases with each additional chronic condition [5] and the presence of activity limitations [2]. As a result of an aging population and increased life expectancy [5], the impact of chronic illness is expected to further grow over time [6].

Chronic conditions are associated with lower levels of subjective well-being, which is defined as a combination of greater positive affect and life satisfaction, and lower negative affect [7]. Decreasing levels of psychological well-being are also associated with increased risk for onset of chronic conditions [8]. For example, people suffering from insomnia report lower levels of subjective well-being [9], and people with asthma, arthritis, diabetes, or heart disease are more likely to indicate that they are dissatisfied or very dissatisfied with their lives [10]. Adolescents with chronic conditions also report poorer emotional well-being than those without chronic conditions [11].

Conversely, among individuals with chronic conditions, high levels of well-being benefit their physical health. Positive affect, eg, may improve survival and recovery rates among people with physical illnesses by activating the autonomic nervous system and the hypothalamic-pituitary-adrenal axis [12]. Positive mood increases the use of self-management strategies among individuals with chronic illness (eg, medication adherence, lifestyle changes, and engaging in preventative behaviors), resulting in fewer complications, symptoms, and activity limitations associated with that illness and, in turn, further boosts positive mood [13]. Greater emotional well-being also predicts better long-term prognoses among patients with physical illnesses [12]; in cardiac patients specifically, an increase of 1 SD in psychological well-being is associated with an 11% reduction in rehospitalization risk [14]. Higher levels of positive affect have demonstrated benefits (eg, better health outcomes and slower disease progression) in a number of other chronic conditions including HIV, chronic heart disease, and coronary artery disease [14-16].

Given the benefits of subjective and psychological well-being among people with chronic conditions, which has been demonstrated in both correlational and experimental studies, a growing body of research has examined the impact of interventions targeting well-being, or mental health, on chronic illness symptoms. A total of 3 key theoretical traditions have been leveraged to improve well-being and mental health in chronic conditions: cognitive behavioral therapy (CBT) [17], mindfulness-based stress reduction (MBSR) [18], and positive psychological interventions (PPIs) [19-21]. In patients living with chronic pain, a robust literature has found that it is possible to enhance subjective well-being, and as a result, to improve pain levels and lessen disability by delivering CBT [22], MBSR [23], or PPIs [24]. Indeed, across a number of chronic somatic diseases, MBSR has a positive impact on mental health and physical health outcomes [25-27], as does CBT [28,29]. Positive psychology approaches differ from CBT and MBSR in that they explicitly target positive affect, which has been associated with lower mortality rates, better treatment compliance, and slower disease progression in a variety of diseases, above and beyond the impact of depression [12,14,30,31]. Interventions that target positive affect have directly improved chronic pain symptoms [24,32,33], even in individuals with more severe disability such as spinal cord injury, neuromuscular disease, or multiple sclerosis [34].

However, there are numerous barriers to accessing in-person interventions, including cost, logistics, and stigma [35,36]. As a result, research has pivoted to explore internet-based interventions, which have the ability to widely increase access to treatment. It is now well established that internet-based behavioral interventions can have a positive effect on psychological well-being [37-39]. However, 1 limitation of previous research is that previous studies targeting well-being tend to draw on either CBT [28,29], mindfulness [23,25,27], or positive psychology [24,32,33], with studies including multiple theoretical approaches comparing those interventions rather than combining them [26]. However, we argue that it is important to embrace all 3 approaches for 2 reasons. First, there is some research to suggest that what works for one patient may not work for all patients; person-activity fit matters. It seems particularly important that the patient believes in the intervention they are using; it must feel authentic to them, and they must also think its premises are plausible [40]. Although mindfulness may be a panacea for one person, to another, it may sound corny. Similarly, CBT may seem overly intellectual to some, whereas it may be just the analytical approach that another patient was looking for. When a patient can choose between multiple approaches, they have the opportunity to select one that they feel is a fit for them, improving their chances of success [41]. Second, we believe that it is important to offer a packaged approach containing multiple frameworks because it seems clear that different psychological interventions operate via different mechanisms (eg, positive affect vs depression) and that intervening through both mechanisms could be better than

intervening via only one [42,43]. To our knowledge, no other study or intervention has combined these methods.

Study Objectives

In this study, we offered a digital intervention platform, Happify, which contains activities that draw from each of the 3 key theoretical approaches. The activities on Happify are adapted efficacious interventions, ie, interventions with evidence from at least two separate research studies, in different samples [44]. Activities are categorized into 5 different groups: savor (activities focusing on mindfulness), thank (activities focusing on gratitude), aspire (activities focusing on optimism, goal setting, and finding meaning and purpose), give (activities focusing on kindness, forgiveness, and prosociality), and empathize (activities focusing on self-compassion and perspective taking). On Happify, activities from the various categories are grouped into tracks, which are designed to focus on a specific issue or problem (eg, reducing stress). Users are able to freely choose a track of interest on the platform and to select between different activity variants in a track. Thus, by completing activities on the Happify platform, users are exposed to well-being interventions from the theoretical traditions of mindfulness, CBT, and positive psychology.

Prior research has demonstrated that using Happify can effectively increase subjective and psychological well-being. Moreover, 1 study of existing Happify users demonstrated that usage was associated with more than a 27% increase in positive emotions over the course of 8 weeks, with greater gains among high-usage participants [45]. In another study using a randomized controlled design, participants randomly assigned to Happify and who completed a minimum of 2 activities per week on average showed statistically greater improvements in depression, anxiety, and resilience compared with a psychoeducation comparison condition or participants with lower platform usage [46]. In addition, a recent study [47] conducted with employees who were experiencing high levels of emotional or workplace distress found that those who used Happify at the recommended level showed greater improvements in resilience than those randomized to a psychoeducational comparison condition or those who did not use their assigned platform. The ideal *dosage* identified in previous internal and published research is 16 activities over the course of 8 weeks [46,47]. Taken together, these results suggest that using Happify can improve well-being in a variety of contexts.

In summary, we argue that improving subjective well-being is important for individuals with chronic conditions because it can help improve their physical condition, thereby reducing the associated costs [1]. We also argue that existing interventions that target well-being can be difficult to access, as they are often offered in person, with associated expenses and other barriers [35,36], and are rarely integrated to contain multiple existing, evidence-based intervention approaches. We provide evidence that a digital platform, Happify, which draws from multiple theoretical traditions to target well-being, has already been established as an efficacious means of improving well-being in both distressed and nondistressed users [45-47]. Although individuals with chronic conditions tend to have lower levels of well-being compared with individuals without chronic conditions [7], there is no reason to believe that they will show a less robust response to these interventions.

In this study, we tested the hypothesis that Happify's efficacy on users without chronic conditions would generalize to a sample of users who report living with a chronic condition. Specifically, we analyzed observational data using Happify to compare the trajectory of change in well-being over time experienced by users on Happify who do and do not report having chronic conditions.

Methods

Recruitment and Sample

Data were drawn from registered users of Happify, a publicly available digital platform that offers games and activities based on research in positive psychology, CBT, and mindfulness. Although Happify is located in the United States, the platform is available worldwide and has been localized in 8 different languages to date. Of the 821 users included in our analyses, the majority used the English language version of the platform (605/821, 73.7% of sample); the remaining users used Happify as translated into German (25/821, 3.1% of sample), Spanish (15/821, 1.8% of sample), Japanese (8/821, 1.0% of sample), French (4/821, 0.5% of sample), Portuguese (3/821, 0.4% of sample), and Chinese (1/821, 0.1% of sample).

When registering with Happify, users provided semipassive consent that their data could be used for research purposes. Specifically, to access Happify content, users were asked to agree to the following statement: "Information that we collect about you also may be combined by us with other information available to us through third parties for research and measurement purposes, including measuring the effectiveness of content, advertising, or programs. This information from other sources may include age, gender, demographic, geographic, personal interests, product purchase activity or other information." Data from all users aged 18 years and older who created accounts on the site between October 29, 2018, and April 4, 2019 (when data were queried), were initially considered; before October 29, 2018, Happify did not ask users about their chronic condition status. Our secondary analysis of Happify consumer data was performed under the supervision of IntegReview, an independent institutional review board.

Materials and Procedures

Screenshots of Happify can be found in a previous publication [46]. After registering with Happify, users completed the onboarding process by responding to a series of questions about their inter- and intrapersonal circumstances, as well as demographic questions such as gender and age, which were collected as a categorical variable with the following options: 18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, and 65 years or older. This was completed to allow for the algorithmic recommendation of a one of a number of *tracks* focused on certain psychosocial topics such as health and well-being, mindfulness and meditation, and relationships. In addition, respondents were asked to select all that apply from a list of chronic conditions, including arthritis, diabetes, insomnia, multiple sclerosis, chronic pain, psoriasis, eczema,

or some other condition. Respondents who selected "some other condition" were not asked to clarify further what that condition was. Finally, the respondent was asked to report on anxiety symptoms by completing the generalized anxiety disorder 2-item (GAD-2) scale [48].

Primary Outcome: Subjective Well-Being

The respondent's subjective well-being was assessed with the Happify Scale, a 9-item measure that includes a positive emotionality component and a life satisfaction component, with higher scores indicating greater well-being [45]. The 4-item subscale measuring positive emotionality was based on the Positive and Negative Affect Schedule [49], a self-report measure that asks participants to indicate the extent to which they experience positive and negative emotions. For example, using a 5-point scale ranging from "Never" to "Very often (almost every day)," participants were asked to respond to the following question, "In the past month, how often have you felt joyous, exuberant, inspired, and/or awestruck?" The 5-item life satisfaction subscale was adapted from the Satisfaction with Life Scale [50] and used to assess satisfaction with different life domains (eg, work, leisure, and relationships). For example, using a 7-point scale ranging from "Very dissatisfied" to "Very satisfied," participants were asked to respond to the following question, "How satisfied do you feel with the relationships in your life?" Scores on the subjective well-being composite range from 0 (low subjective well-being) to 100 (high subjective well-being). Scale validation using a general population sample from Amazon MTurk indicated that scores between 46 and 49 corresponded to the 25th percentile, scores between 61 and 63 corresponded to the 50th percentile, and scores between 75 and 77 corresponded to the 75th percentile of the Happify Scale. Internal validation data indicated that composite scale scores had acceptable reliability (alpha=.89) and were significantly and strongly associated with both subjective happiness [51] at r=0.78 and a measure of depressive symptoms (Center for Epidemiologic Studies Depression Scale [52]) at r=−0.75.

Participants were prompted to complete the Happify Scale on the day after completing the platform registration onboarding process and every 2 weeks thereafter. In each case, the assessment was optional, and users were able to exit out of the assessment without completing it if they wished. As a result, there was considerable variability in terms of how many assessments users completed and when those assessments were completed. For each individual, we calculated an average time between any 2 assessments (in days). The average of this average across the sample is 30.69 days (SD=21.10), ranging from 11.83 to 149.

Statistical Analysis

Descriptive statistics were stratified by self-reported chronic condition status (yes vs no). Group differences in baseline variables were examined using chi-square tests for categorical characteristics and *t* tests or Mann Whitney *U* tests (in the case of non-normally distributed variables) for continuous variables.

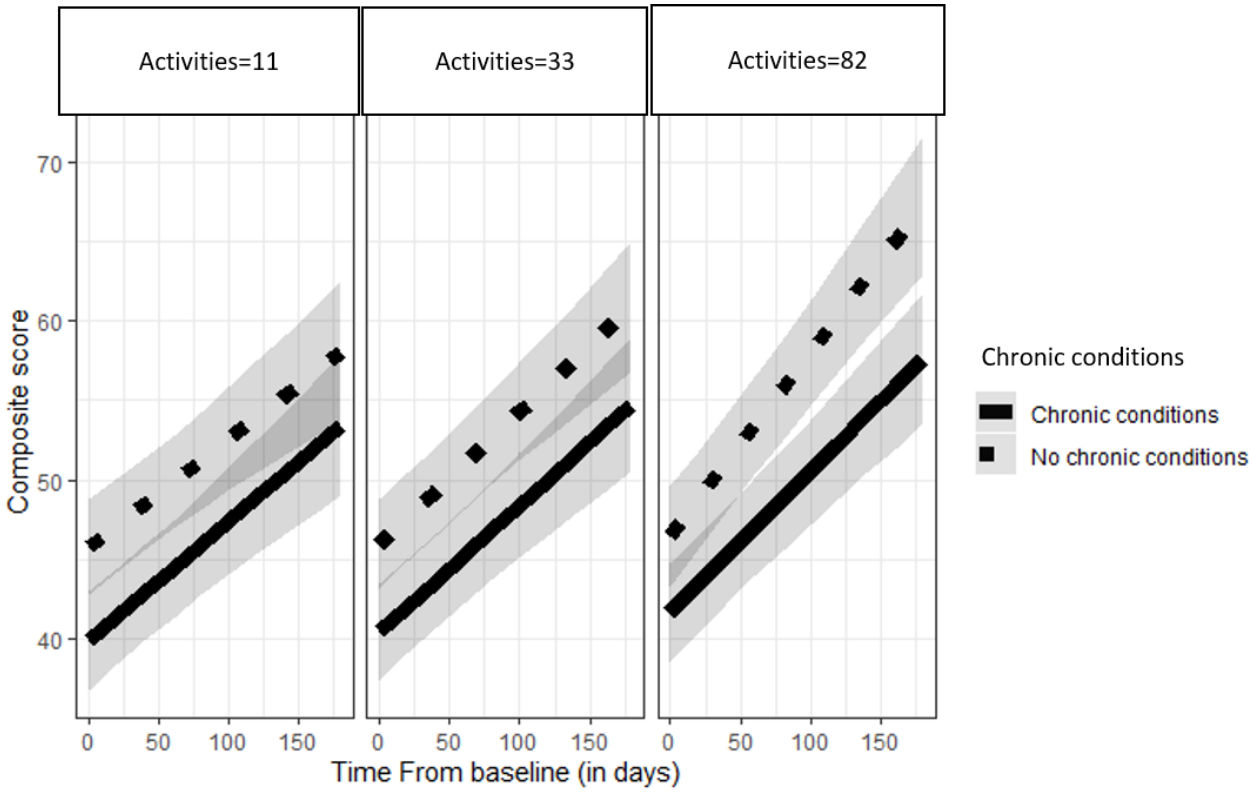
To evaluate changes in well-being over time, a mixed effects linear regression model was fit for subjective well-being. The predictor variable of key interest was self-report of any of the 8 chronic conditions gathered at baseline. A binary variable was created to indicate having 1 or more chronic conditions vs none. The following covariates were included as control variables: gender, age category, number of activities completed on Happify, baseline anxiety [48], and time from baseline to each assessment (in days). All of these were treated as fixed effects.

Normally distributed person-specific random effects were included to account for varying numbers of follow-up assessments. To test whether changes in outcome measures differed between those reporting a chronic condition and those not reporting a chronic condition, all interaction terms between time from baseline, number of activities completed, and chronic condition status were included. Adherence to modeling assumptions was tested using residual plots (eg, Q-Q plots to examine if residuals followed a Gaussian distribution) and was met.

All computations were done in R, version 3.6.1 [53]. All linear mixed models were fitted using the R packages *lme4* [54] and *lmerTest* [55]. *P* values were calculated from Satterthwaite approximations for degrees of freedom [56]. All tests were 2-sided, and *P* values less than .05 were considered statistically significant.

For Figure 1, data points are derived from predicted values of linear mixed models. Error bars are based on 95% CIs from those predicted values.

Figure 1. Change in well-being over time for users with and without a chronic condition. Facets are broken into the 25th, 50th, and 75th percentile of activities completed. Gray bands around the chronic condition lines reflect 95% CIs.



Data Exclusion

During the study period, 6801 new users created accounts and completed the baseline assessment. Of these, data were excluded from those users who never completed any activities (n=1931) or who did not complete their self-report measures in a way that makes logical sense in relation to their usage of the platform (n=1058). Specifically, to be included, they were required to use Happify within 30 days of taking their initial baseline assessment—otherwise, their baseline assessment may or may not have accurately represented their state when they started using Happify. They were also required to have taken their final assessment within 30 days of their final activity to maximize chances that self-report scores were representative of the user's psychological state when usage was terminated. In addition, users were excluded if they did not receive a minimum of 6 weeks' exposure to Happify. They were not required to use Happify at any particular level during that time, but they were required to at least have had access to Happify for 6 weeks or more (2939 users were excluded by this criterion). Finally, 52 participants had missing onboarding questions because of a server error and were excluded. The final sample consisted of 821 users who had access to the platform between 42 and 179 days.

Power

A statistical power analysis was performed for sample size estimation based on data from a randomized study examining Happify's efficacy [46]. The effect size (ES) in this study, based on participants randomly assigned to Happify (vs a control condition) and who used the platforms at the recommended level (ie, completing an average of 2 activities per week), was η^2=0.021, classified as small by Cohen [57] criteria. Specifying alpha=.05 and power=0.80, the projected sample size needed with this ES (GPower 3.1.9.2 [58]) for a repeated-measures analysis with a between-within interaction was n=94. This calculation was made with the conservative estimate of only 2 assessments and a correlation between repeated measures of *r*=0.50. Thus, this study was adequately powered to detect the effect of interest.

Results

Baseline Sample Characteristics

Table 1 displays the baseline sample characteristics of individuals who reported having a chronic condition and those who did not. Most importantly, the 2 groups were not statistically different with respect to the number of activities completed or the number of assessments finished. There were significant differences, however, between the 2 groups in terms of age, such that users reporting a chronic condition were more likely to fall into the older age categories and less likely to fall in the 18 to 24-year-old category. All users with a chronic condition reported having a health condition impacting their well-being, although 25 individuals without a chronic condition also reported this. Users with chronic conditions were also significantly more likely to be characterized as having anxiety based on their GAD-2 scores (65.6% vs 55.0%).

Table 1. Baseline sample characteristics and platform usage of the users with and without chronic conditions.

Characteristic	Chronic conditions (n=450)	No chronic conditions (n=371)	*P* value
Female, n (%)	372 (82.7)	305 (82.2)	.94
Age (years), n (%)			<.001
18-24	73 (16.2)	103 (27.8)	
25-34	139 (30.9)	116 (31.3)	
35-44	109 (24.2)	88 (23.7)	
45-54	86 (19.1)	41 (11.1)	
55-64	35 (7.8)	20 (5.4)	
65+	8 (1.8)	3 (0.8)	
Activities completed, mean (SD)	56.20 (86.63)	50.50 (64.51)	.29
Total time elapsed between first and last activity (in days), mean (SD)	72.19 (27.19)	76.23 (30.79)	.046
Total number of assessments, mean (SD)	4.18 (1.76)	4.20 (1.69)	.86
Is there a health condition or concern that impacts your happiness or well-being currently?, n (%)			<.001
Not at all	0 (0.0)	346 (93.3)	
Yes, not major	259 (57.6)	23 (6.2)	
Yes, very much	191 (42.4)	2 (0.5)	
Arthritis, n (%)	31 (8.5)	0 (0.0)	<.001
Chronic pain, n (%)	75 (20.5)	0 (0.0)	<.001
Insomnia, n (%)	92 (25.2)	0 (0.0)	<.001
Multiple sclerosis, n (%)	3 (0.8)	0 (0.0)	.33
Psoriasis, n (%)	7 (1.9)	0 (0.0)	.05
Diabetes, n (%)	13 (3.6)	0 (0.0)	.003
Eczema, n (%)	0 (0.0)	0 (0.0)	Not applicable
Other chronic condition, n (%)	269 (73.7)	0 (0.0)	<.001
Number of chronic conditions, median (IQR)	1.00 (1.00-2.00)	0.00 (0.00-0.00)	<.001[a]
Positive emotion score, mean (SD)	32.34 (17.65)	36.43 (20.52)	.002
Life satisfaction score, mean (SD)	44.78 (21.55)	51.36 (22.55)	<.001
Subjective well-being, mean (SD)	38.34 (17.40)	43.65 (19.13)	<.001
Generalized anxiety disorder 2-item scores, median (IQR)	4.00 (2.00-6.00)	3.00 (2.00-5.00)	.002

[a]Mann Whitney *U* test.

The 2 groups were significantly different at baseline in terms of positive emotionality, as users with a chronic condition (mean 32.34, SD 17.65) had lower positive emotion scores than users without a chronic condition (mean 36.43, SD 0.52); life satisfaction baseline scores were similarly statistically different for users with a chronic condition (mean 44.78, SD 21.55) and those without a chronic condition (mean 51.36, SD 22.55), with users with a chronic condition scoring lower. Users with a chronic condition also had lower scores on the composite subjective well-being scale (mean 38.34, SD 17.40) than users without a chronic condition (mean 43.65, SD 19.13).

For those users with a chronic condition, the most common reported category was “other,” followed by insomnia, chronic pain, and arthritis. The most common number of reported conditions was 1; however, 136 users (136/450, 30.2% of the users with chronic conditions) reported having 2 or more.

Change in Well-Being Over Time

For subjective well-being scores at final assessment, there were main effects for chronic condition status (b=4.82; SE=1.51; P<.01) and baseline GAD-2 score (b=−3.43; SE=0.30; P<.01). Users reporting a chronic condition and users reporting higher levels of anxiety had lower subjective well-being scores at their final assessment. In addition, there was an effect for time from baseline (b=0.071; SE=0.010; P<.01) and number of activities completed (b=0.03; SE=0.009; P<.01). Higher subjective well-being scores occurred among users who had been active users on Happify for longer and who had completed higher numbers of activities. Finally, there was a 2-way interaction between number of activities completed and time from baseline

(b=0.0002; SE=0.00006; P<.01). For all users, completing more activities and doing so over increasingly longer periods interacted to produce improved well-being scores. There were no other significant interactions. However, there was no significant interaction between time from baseline and chronic condition status (b=−0.013; SE=0.071; P=.46). These results indicate that users with and without a chronic condition both experienced equal well-being improvements from using Happify and from completing higher numbers of activities on the platform; both groups of users showed the same pattern of change in well-being over time. Depictions of changes in the subjective well-being scale across time for both users with and without a chronic condition are presented in Figure 1. Level of activities completed is split into 3 facets for the 25th (activities completed=11), 50th (activities completed=33), and 75th (activities completed=82) percentiles.

Discussion

Principal Findings

A key objective of this study was to explore whether a digital intervention could reliably improve subjective well-being among users living with a chronic condition. We were particularly interested in testing the impact of an intervention that targets subjective well-being because of the demonstrated benefits of subjective well-being, and especially positive affect, among individuals with chronic conditions such as greater self-management of their condition [13], better long-term prognoses [24], and better health outcomes [7,12,14-16]. Although other research suggests that CBT, MBSR, and PPIs increase subjective well-being and improve physical health outcomes among people with chronic conditions [22-24], many of these interventions are in person, making scaling difficult. In addition, these other interventions draw on just a single theoretical approach [26], rather than combining strengths from all 3 approaches into a single intervention. We explored the impact of improving well-being on users with chronic conditions using observational data from Happify, an existing commercial platform that integrates principles from CBT, mindfulness, and positive psychology and contains users both with and without chronic conditions. Although prior research supports the idea that Happify improves well-being among physically healthy users [35,45-47], no research to date has tested Happify's efficacy in users who are dealing with a chronic disease.

Consistent with other studies demonstrating the effectiveness of PPIs [32-34], mindfulness [23,26], and CBT [26] on subjective well-being among people with chronic conditions, data from this study support the conclusion that users with a chronic condition experienced significant improvement in subjective well-being over time, and their change trajectory did not differ from those without a chronic condition. Users who completed more activities over a longer period showed the greatest amount of improvement, a finding that is consistent with past research [45-47], and chronic condition status did not change this result. In other words, the presence of a chronic stressor, at least in the case of chronic conditions, does not appear to prevent users from experiencing improvements in well-being when they use Happify.

Although previous research typically focused on specific conditions such as chronic pain [32], spinal cord injury, multiple sclerosis, neuromuscular disease [34], or osteoarthritis [33], we used a noncategorical approach [59], grouping all users who self-reported a chronic condition together. Consequently, data from this study do not speak to whether effectiveness differed by type of condition. However, other studies found no differences in subjective well-being based on type of condition or visibility of the condition [11]. In addition, researchers have advocated for the use of a noncategorical approach in applied research on chronic conditions because people with chronic conditions share common problems that go above and beyond the specific symptoms of their particular illness [59], including lower subjective well-being. This approach may be particularly relevant when evaluating interventions as communities may have a large proportion of individuals with chronic conditions but only small numbers of people with specific conditions [59]. Arguably, then, by using a noncategorical approach and including participants with a variety of chronic conditions, our findings have more direct applicability to those suffering from these conditions.

Future Directions

This study provides preliminary evidence that Happify can significantly improve subjective well-being among people with chronic conditions, despite the fact that people with chronic conditions also are more likely to suffer from more serious psychological distress. For example, although depression has a prevalence of 10% to 20% in the general population, among individuals living with a chronic condition, depression rates range from 35% to 50% [60-63]. Individuals with chronic pain are 4 times more likely to have either a depressive or anxiety disorder than those without chronic pain [64], and the incidence of comorbid depression and anxiety disorders is greater than independent diagnoses of either depression or anxiety [65]. People with insomnia are 2 times more likely to develop depression than those without insomniac symptoms [66], and chronic insomnia is associated with an elevated risk for anxiety disorders [67].

Importantly, the burden of chronic illness can be amplified when poor mental health, especially depression, is also present. Chronic pain patients with comorbid depression and anxiety report greater pain severity and pain-related disability as well as poorer health-related quality of life than people with pain alone [68]. Depression also predicts poor treatment adherence, greater frequency of complications, and higher mortality rates among people with diabetes [69,70]. In fact, in a study of 60 countries, respondents with a chronic physical condition and comorbid depression had the worst health scores overall [71]. Although previous research has shown that Happify users report fewer depressive and anxiety symptoms after 8 weeks [46], we only assessed subjective well-being in this study. Therefore, in future research, it will be important to determine whether Happify also helps to reduce depressive and anxiety symptoms among people with chronic conditions.

Another important direction for future research is to explore how, specifically, Happify usage helps to improve mental well-being. Previous research suggests that mindfulness

programs have been effective in reducing depressive symptoms among individuals with chronic pain by reducing pain catastrophizing and psychological distress [72]. Other studies show that Web-based interventions lead to fewer activity limitations [73] and improved pain acceptance [23]. Future research should include measures assessing participants' physical condition as well, including activity limitations, pain severity and acceptance, and pain catastrophizing, to determine whether Happify usage also impacts these outcomes and whether they mediate the relationship between Happify use and improved subjective well-being. Similarly, research exploring the long-term benefits of Happify on users' psychological and physical well-being would be valuable to determine whether it, like some other interventions [74], can also help to lower costs associated with chronic conditions by reducing health care usage.

Limitations and Strengths

This study was a naturalistic, observational study of existing Happify users. Although observational studies are an important tool in the assessment of health-related outcomes [75], there are also several limitations associated with the lack of control in these designs. One limitation is that although user data were collected in a realistic context, participants were all Happify registrants who made their way to the platform naturally and, consequently, may differ from people in the general population that do not use Happify. Specifically, this study and several others reliably find that users on Happify are more distressed than the general population. However, the sample is likely to be biased in the same direction as future Happify participants, so any conclusions drawn about this sample may well be applicable to our population of interest—ie, those who use Happify in the future.

Usage patterns observed in this study were also naturally occurring, as compared with those that may be observed in a more controlled study with participation incentives and more frequent (potentially annoying or invasive) reminders. Nevertheless, given that the Happify platform tested in this study is a commercial product, freely available to the public, and just as easy to quit as it is to sign up, dropout levels were higher than would be observed in a more controlled setting. The resulting sample consisted of only the most dedicated users. Therefore, self-selection is a concern for this type of study design. However, even randomized controlled trials (RCTs) can suffer from this, as unmeasured moderating variables may influence a participant's willingness to participate in a randomized study [57].

Finally, because this study is observational, we cannot rule out the possibility that the users with a chronic condition were different from those without a chronic condition in ways we did not measure; as chronic condition status cannot be randomized, we did not have the benefit of randomization to address systematic biases. Moreover, although we collected respondent data on a number of chronic conditions, our list of chronic conditions was not exhaustive. Approximately 73.7% (605/821) of respondents in the chronic condition group self-identified as having "some other condition" for which we have no additional information. It may also be that users with and without a chronic condition differ on meaningful but unmeasured covariates. For example, as Happify is a commercial product and the analyses included in this study are secondary analyses from Happify's consumer base and not a randomized clinical trial, we did not have access to user information that might be relevant here, such as access to other mental health-related treatments. Such differences between groups or omitted variables can contribute to biased estimates of treatment effects [76]. In addition, users in this study completed activities and assessments at varying times and to a varying degree. This also creates difficulty in assessing change over time compared with an RCT, where the intervention and assessments are planned and given at regular intervals. However, linear mixed models have been shown to be effective at controlling for such unbalanced data occurring at varying time points [77].

Conclusions

It is all too easy in the world of digital well-being interventions, the use of which is largely unregulated, to assume that an intervention that works in one population can safely be generalized to other populations. We would argue that it is important to understand who may be in the sample of consumers interacting with a digital intervention and to evaluate whether there are subgroups of users for whom the intervention fails to produce results. Although in the case of this paper, we were able to ascertain that Happify's effects on well-being do not differ significantly between users with chronic conditions and those without chronic conditions, we could also have found that users with chronic conditions need something else; only by evaluating subgroup data can we gain confidence in our ability to generalize, as a freely available digital intervention inevitably will. In summary, this study provides valuable observational evidence of the efficacy of Happify's use among real users living with chronic conditions under naturalistic conditions. Given these data, future research should seek to replicate these effects under more controlled conditions, such as RCTs, and explore the impact of Happify's use on other important outcomes associated with chronic conditions such as depressive and anxiety symptoms as well as physical and health-related outcomes.

Authors' Contributions

AP contributed substantially to the study aims and scope, wrote the initial draft of the Introduction and Discussion, and oversaw the writing team. AW contributed substantially to the revision of all sections and to the drafting of the Methods section, and also provided support in responding to review comments on methodology and drafted the revision response. GK provided substantive support in the drafting of the Introduction and Discussion sections. JS provided substantive support in the drafting of the Introduction

and Discussion sections as well as general manuscript support. EB led the revision of the Introduction and Discussion sections in response to reviewer feedback. RH crafted the statistical approach and drafted and revised the Results section.

Conflicts of Interest

ACP, ALW, GMK, EMB, and JLS are employees of Happify. RDH is a paid consultant with Happify.

References

1. National Center for Chronic Disease Prevention and Health Promotion. Centers for Disease Control and Prevention. Atlanta, GA; 2009. The Power of Prevention: Chronic Disease...The Public Health Challenge of the 21st Century. URL: https://www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf [accessed 2019-12-19]
2. Anderson G. Robert Wood Johnson Foundation. 2010 Jan 1. Chronic Care: Making the case for ongoing care. URL: https://www.rwjf.org/en/library/research/2010/01/chronic-care.html [accessed 2019-12-19]
3. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003 Nov 10;163(20):2433-2445. [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
4. Collins JJ, Baase CM, Sharda CE, Ozminkowski RJ, Nicholson S, Billotti GM, et al. The assessment of chronic health conditions on work performance, absence, and total economic impact for employers. J Occup Environ Med 2005 Jun;47(6):547-557. [doi: 10.1097/01.jom.0000166864.58664.29] [Medline: 15951714]
5. Anderson G, Horvath J. The growing burden of chronic disease in America. Public Health Rep 2004;119(3):263-270 [FREE Full text] [doi: 10.1016/j.phr.2004.04.005] [Medline: 15158105]
6. Tackling the burden of chronic diseases in the USA. Lancet 2009 Jan 17;373(9659):185. [doi: 10.1016/S0140-6736(09)60048-9] [Medline: 19150686]
7. Diener E, Scollon CN, Lucas RE. The evolving concept of subjective well-being: the multifaceted nature of happiness. In: Diener E, editor. Assessing Well-Being. Dordrecht: Springer; 2009:67-100.
8. Ryff CD, Radler BT, Friedman EM. Persistent psychological well-being predicts improved self-rated health over 9-10 years: longitudinal evidence from MIDUS. Health Psychol Open 2015;2(2):pii: 2055102915601582 [FREE Full text] [doi: 10.1177/2055102915601582] [Medline: 26617988]
9. Hamilton NA, Gallagher MW, Preacher KJ, Stevens N, Nelson CA, Karlson C, et al. Insomnia and well-being. J Consult Clin Psychol 2007 Dec;75(6):939-946. [doi: 10.1037/0022-006X.75.6.939] [Medline: 18085910]
10. Strine TW, Chapman DP, Balluz LS, Moriarty DG, Mokdad AH. The associations between life satisfaction and health-related quality of life, chronic illness, and health behaviors among US community-dwelling adults. J Community Health 2008 Feb;33(1):40-50. [doi: 10.1007/s10900-007-9066-4] [Medline: 18080207]
11. Wolman C, Resnick MD, Harris LJ, Blum RW. Emotional well-being among adolescents with and without chronic conditions. J Adolesc Health 1994 May;15(3):199-204. [doi: 10.1016/1054-139x(94)90504-5] [Medline: 8075089]
12. Lamers SM, Bolier L, Westerhof GJ, Smit F, Bohlmeijer ET. The impact of emotional well-being on long-term recovery and survival in physical illness: a meta-analysis. J Behav Med 2012 Oct;35(5):538-547 [FREE Full text] [doi: 10.1007/s10865-011-9379-8] [Medline: 21918889]
13. de Ridder D, Geenen R, Kuijer R, van Middendorp H. Psychological adjustment to chronic disease. Lancet 2008 Jul 19;372(9634):246-255. [doi: 10.1016/S0140-6736(08)61078-8] [Medline: 18640461]
14. DuBois CM, Lopez OV, Beale EE, Healy BC, Boehm JK, Huffman JC. Relationships between positive psychological constructs and health outcomes in patients with cardiovascular disease: A systematic review. Int J Cardiol 2015 Sep 15;195:265-280 [FREE Full text] [doi: 10.1016/j.ijcard.2015.05.121] [Medline: 26048390]
15. Ironson G, Hayward H. Do positive psychosocial factors predict disease progression in HIV-1? A review of the evidence. Psychosom Med 2008 Jun;70(5):546-554 [FREE Full text] [doi: 10.1097/PSY.0b013e318177216c] [Medline: 18541905]
16. Hevey D, McGee HM, Horgan JH. Comparative optimism among patients with coronary heart disease (CHD) is associated with fewer adverse clinical events 12 months later. J Behav Med 2014 Apr;37(2):300-307. [doi: 10.1007/s10865-012-9487-0] [Medline: 23274764]
17. Beck AT. Cognitive Therapy of Depression. New York: The Guilford Press; 1979.
18. Praissman S. Mindfulness-based stress reduction: a literature review and clinician's guide. J Am Acad Nurse Pract 2008 Apr;20(4):212-216. [doi: 10.1111/j.1745-7599.2008.00306.x] [Medline: 18387018]
19. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
20. Schueller SM, Parks AC. Disseminating self-help: positive psychology exercises in an online trial. J Med Internet Res 2012 Jun 25;14(3):e63 [FREE Full text] [doi: 10.2196/jmir.1850] [Medline: 22732765]
21. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
22. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]

23. Dowd H, Hogan MJ, McGuire BE, Davis MC, Sarma KM, Fish RA, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation: a randomized controlled study. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]
24. Hausmann LR, Parks A, Youk AO, Kwoh CK. Reduction of bodily pain in response to an online positive activities intervention. J Pain 2014 May;15(5):560-567. [doi: 10.1016/j.jpain.2014.02.004] [Medline: 24568751]
25. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: a meta-analysis. J Psychosom Res 2010 Jun;68(6):539-544. [doi: 10.1016/j.jpsychores.2009.10.005] [Medline: 20488270]
26. Tovote KA, Fleer J, Snippe E, Peeters AC, Emmelkamp PM, Sanderman R, et al. Individual mindfulness-based cognitive therapy and cognitive behavior therapy for treating depressive symptoms in patients with diabetes: results of a randomized controlled trial. Diabetes Care 2014 Sep;37(9):2427-2434. [doi: 10.2337/dc13-2918] [Medline: 24898301]
27. Gregg JA, Callaghan GM, Hayes SC, Glenn-Lawson JL. Improving diabetes self-management through acceptance, mindfulness, and values: a randomized controlled trial. J Consult Clin Psychol 2007 Apr;75(2):336-343. [doi: 10.1037/0022-006X.75.2.336] [Medline: 17469891]
28. Sivertsen B, Omvik S, Pallesen S, Bjorvatn B, Havik OE, Kvale G, et al. Cognitive behavioral therapy vs zopiclone for treatment of chronic primary insomnia in older adults: a randomized controlled trial. J Am Med Assoc 2006 Jun 28;295(24):2851-2858. [doi: 10.1001/jama.295.24.2851] [Medline: 16804151]
29. Boschen KA, Robinson E, Campbell KA, Muir S, Oey E, Janes K, et al. Results from 10 years of a CBT pain self-management outpatient program for complex chronic conditions. Pain Res Manag 2016:4678083 [FREE Full text] [doi: 10.1155/2016/4678083] [Medline: 27891062]
30. Moskowitz JT, Epel ES, Acree M. Positive affect uniquely predicts lower risk of mortality in people with diabetes. Health Psychol 2008 Jan;27(1S):S73-S82. [doi: 10.1037/0278-6133.27.1.S73] [Medline: 18248108]
31. Bassett SM, Schuette SA, O'Dwyer LC, Moskowitz JT. Positive affect and medication adherence in chronic conditions: a systematic review. Health Psychol 2019 Nov;38(11):960-974. [doi: 10.1037/hea0000778] [Medline: 31368717]
32. Boselie JJ, Vancleef LM, Peters ML. Filling the glass: effects of a positive psychology intervention on executive task performance in chronic pain patients. Eur J Pain 2018 Aug;22(7):1268-1280 [FREE Full text] [doi: 10.1002/ejp.1214] [Medline: 29573513]
33. Hausmann LR, Youk A, Kwoh CK, Ibrahim SA, Hannon MJ, Weiner DK, et al. Testing a positive psychological intervention for osteoarthritis. Pain Med 2017 Oct 1;18(10):1908-1920 [FREE Full text] [doi: 10.1093/pm/pnx141] [Medline: 29044408]
34. Müller R, Gertz KJ, Molton IR, Terrill AL, Bombardier CH, Ehde DM, et al. Effects of a tailored positive psychology intervention on well-being and pain in individuals with chronic pain and a physical disability: a feasibility trial. Clin J Pain 2016 Jan;32(1):32-44. [doi: 10.1097/AJP.0000000000000225] [Medline: 25724020]
35. Muñoz RF, Bunge EL, Chen K, Schueller SM, Bravin JI, Shaughnessy EA, et al. Massive open online interventions. Clin Psychol Sci 2015;4(2):194-205. [doi: 10.1177/2167702615583840]
36. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]
37. 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 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
38. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
39. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
40. Schueller SM. Preferences for positive psychology exercises. J Pos Psychol 2010;5(3):192-203. [doi: 10.1080/17439761003790948]
41. Sheldon KM, Elliot AJ. Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model. J Pers Soc Psychol 1999 Mar;76(3):482-497. [doi: 10.1037//0022-3514.76.3.482] [Medline: 10101878]
42. Lemmens LH, Müller VN, Arntz A, Huibers MJ. Mechanisms of change in psychotherapy for depression: an empirical update and evaluation of research aimed at identifying psychological mediators. Clin Psychol Rev 2016 Dec;50:95-107. [doi: 10.1016/j.cpr.2016.09.004] [Medline: 27770716]
43. Desrosiers A, Vine V, Klemanski DH, Nolen-Hoeksema S. Mindfulness and emotion regulation in depression and anxiety: common and distinct mechanisms of action. Depress Anxiety 2013 Jul;30(7):654-661 [FREE Full text] [doi: 10.1002/da.22124] [Medline: 23592556]
44. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998 Feb;66(1):7-18. [doi: 10.1037//0022-006x.66.1.7] [Medline: 9489259]

45. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Correction: seeing the 'big' picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data. J Med Internet Res 2017 Dec 19;19(12):e347 [FREE Full text] [doi: 10.2196/jmir.8099] [Medline: 29258059]
46. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. Int J Wellbeing 2018;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]
47. Williams AL, Parks AC, Cormier G, Stafford J, Whillans A. Improving resilience among employees high in depression, anxiety, and workplace distress. Int J Manag Res 2018 Dec;9(1-2):4-22 [FREE Full text]
48. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
49. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
50. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
51. Lyubomirsky S, Lepper HS. A measure of subjective happiness: preliminary reliability and construct validation. Soc Indic Res 1999;46(2):137-155 [FREE Full text] [doi: 10.1023/A:1006824100041]
52. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
53. R: The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing URL: http://www.R-project.org/ [accessed 2019-12-19]
54. Bates D, Mächler M, Bolker BS, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
55. Kuznetsova A, Brockhoff PB, Christensen RH. LmerTest Package: tests in linear mixed effects models. J Stat Softw 2017;82(13):1-26 [FREE Full text] [doi: 10.18637/jss.v082.i13]
56. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
57. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Routledge; 1988.
58. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
59. Stein RE, Jessop DJ. What diagnosis does not tell: the case for a noncategorical approach to chronic illness in childhood. Soc Sci Med 1989;29(6):769-778. [doi: 10.1016/0277-9536(89)90157-3] [Medline: 2772671]
60. Miller LR, Cano A. Comorbid chronic pain and depression: who is at risk? J Pain 2009 Jun;10(6):619-627. [doi: 10.1016/j.jpain.2008.12.007] [Medline: 19398383]
61. Feinstein A. Multiple sclerosis and depression. Mult Scler 2011 Nov;17(11):1276-1281. [doi: 10.1177/1352458511417835] [Medline: 22058085]
62. Mancuso CA, Peterson MG, Charlson ME. Effects of depressive symptoms on health-related quality of life in asthma patients. J Gen Intern Med 2000 May;15(5):301-310 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.07006.x] [Medline: 10840265]
63. Peyrot M, Rubin RR. Levels and risks of depression and anxiety symptomatology among diabetic adults. Diabetes Care 1997 Apr;20(4):585-590. [doi: 10.2337/diacare.20.4.585] [Medline: 9096984]
64. Gureje O, von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. J Am Med Assoc 1998 Jul 8;280(2):147-151. [doi: 10.1001/jama.280.2.147] [Medline: 9669787]
65. Scott KM, Bruffaerts R, Tsang A, Ormel J, Alonso J, Angermeyer MC, et al. Depression-anxiety relationships with chronic physical conditions: results from the World Mental Health Surveys. J Affect Disord 2007 Nov;103(1-3):113-120. [doi: 10.1016/j.jad.2007.01.015] [Medline: 17292480]
66. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
67. Neckelmann D, Mykletun A, Dahl AA. Chronic insomnia as a risk factor for developing anxiety and depression. Sleep 2007 Jul;30(7):873-880 [FREE Full text] [doi: 10.1093/sleep/30.7.873] [Medline: 17682658]
68. Bair MJ, Wu J, Damush TM, Sutherland JM, Kroenke K. Association of depression and anxiety alone and in combination with chronic musculoskeletal pain in primary care patients. Psychosom Med 2008 Oct;70(8):890-897 [FREE Full text] [doi: 10.1097/PSY.0b013e318185c510] [Medline: 18799425]
69. Egede LE, Ellis C. Diabetes and depression: global perspectives. Diabetes Res Clin Pract 2010 Mar;87(3):302-312. [doi: 10.1016/j.diabres.2010.01.024] [Medline: 20181405]
70. Gonzalez JS, Kane NS, Binko DH, Shapira A, Hoogendoorn CJ. Tangled up in blue: unraveling the links between emotional distress and treatment adherence in type 2 diabetes. Diabetes Care 2016 Dec;39(12):2182-2189 [FREE Full text] [doi: 10.2337/dc16-1657] [Medline: 27797932]

71. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007 Sep 8;370(9590):851-858. [doi: 10.1016/S0140-6736(07)61415-9] [Medline: 17826170]
72. Brooks JM, Iwanaga K, Cotton BP, Deiches J, Blake J, Chiu C, et al. Perceived mindfulness and depressive symptoms among people with chronic pain. J Rehabil 2018;84(2):33-39 [FREE Full text] [Medline: 30686840]
73. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. Pain 2009 Nov;146(1-2):205-213 [FREE Full text] [doi: 10.1016/j.pain.2009.07.034] [Medline: 19695776]
74. Turner JA. Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. J Consult Clin Psychol 1982 Oct;50(5):757-765. [doi: 10.1037//0022-006x.50.5.757] [Medline: 6216275]
75. Black N. Why we need observational studies to evaluate the effectiveness of health care. Br Med J 1996 May 11;312(7040):1215-1218 [FREE Full text] [doi: 10.1136/bmj.312.7040.1215] [Medline: 8634569]
76. D'Agostino Jr RB, D'Agostino Sr RB. Estimating treatment effects using observational data. J Am Med Assoc 2007 Jan 17;297(3):314-316. [doi: 10.1001/jama.297.3.314] [Medline: 17227985]
77. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat Med 1997 Oct 30;16(20):2349-2380. [doi: 10.1002/(sici)1097-0258(19971030)16:20<2349::aid-sim667>3.0.co;2-e] [Medline: 9351170]

Abbreviations

CBT: cognitive behavioral therapy
ES: effect size
GAD-2: generalized anxiety disorder 2-item
MBSR: mindfulness-based stress reduction
PPIs: positive psychological interventions
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 10.09.19; peer-reviewed by S Bassett, S Schuette, O Rivera, N Kennedy, M Khazaee-Pool; comments to author 03.10.19; revised version received 01.11.19; accepted 10.12.19; published 10.01.20.

Please cite as:
Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD
The Effects of a Digital Well-Being Intervention on Patients With Chronic Conditions: Observational Study
J Med Internet Res 2020;22(1):e16211
URL: https://www.jmir.org/2020/1/e16211
doi:10.2196/16211
PMID:31922491

Original Paper

A Culturally Targeted eLearning Module on Organ Donation (Promotoras de Donación): Design and Development

Gerard P Alolod[1*], MPhil; Heather Gardiner[2*], PhD, MPH; Chidera Agu[2], MPH, MSW; Jennie L Turner[2], MS; Patrick J Kelly[2], BSPH; Laura A Siminoff[1], PhD; Elisa J Gordon[3], PhD, MPH; Robert Norden[4], MA; Theresa A Daly[4], MS; Amanda Benitez[5], MPH; Ilda Hernandez[5], CHW; Nancy Guinansaca[6]; Lori Ramos Winther[6], MA, MPH; Caroline D Bergeron[7], DrPH; Antonette Montalvo[8], MSN; Tony Gonzalez[8], MDiv

[1]College of Public Health, Temple University, Philadelphia, PA, United States

[2]Department of Social and Behavioral Sciences, College of Public Health, Temple University, Philadelphia, PA, United States

[3]Division of Transplantation, Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[4]Gift of Life Institute, Philadelphia, PA, United States

[5]Enlace Chicago, Chicago, IL, United States

[6]Centro San Bonifacio, Chicago, IL, United States

[7]Bexar County Community Health Collaborative, San Antonio, TX, United States

[8]Esperanza Health Center, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Gerard P Alolod, MPhil
College of Public Health
Temple University
1700 N Broad St
Suite 417
Philadelphia, PA, 19121
United States
Phone: 1 2152049186
Email: gpalolod@temple.edu

Abstract

Background: As an overrepresented population on the transplant waitlist, stagnated rates of organ donation registration among Latinxs must be redressed. Promotoras (community health workers), who are effective at advocating and spearheading health promotion efforts in the Latinx community, show promise in their ability to educate about organ donation and donor registration.

Objective: This study aimed (1) to develop an interactive, evidence-based program to educate promotoras about organ donation, the need for organ donors in the Latinx American community, and ways to register as deceased organ donors and (2) to train promotoras to lead discussions about organ donation and to promote the act of donor registration.

Methods: In partnership with 4 promotoras organizations, the culturally targeted *Promotoras de Donación* eLearning module was developed based on input from 12 focus groups conducted with Latina women (n=61) and promotoras (n=37). Formative work, existing literature, the Vested Interest Theory, and the Organ Donation Model guided curriculum development. In partnership with the Gift of Life Institute and regional promotoras, the curriculum was designed, filmed, and developed in a visually appealing module interface. The module was beta-tested with promotoras before launch.

Results: *Promotoras de Donación*, available in Spanish with English subtitling, lasts just over an hour. The module comprised 6 sections including various activities and videos, with the curriculum divided into a skills-based communication component and a didactic educational component. Pre- and posttests assessed the module's direct effects on promotoras' organ donation knowledge and attitudes as well as confidence promoting the act of donor registration.

Conclusions: This novel, theoretically and empirically based intervention leveraged the existing network of promotoras to promote the act of donor registration. Future research should assess whether the module helps increase rates of donor registration within Latinx communities and reduce disparities in access to transplantation.

Trial Registration: ClinicalTrials.gov NCT04007419; https://www.clinicaltrials.gov/ct2/show/NCT04007419

(J Med Internet Res 2020;22(1):e15793) doi:10.2196/15793

KEYWORDS

Hispanic Americans; organ donation; program development; program evaluation; education

Introduction

Background

Although the number of deceased organ donors in the United States has increased modestly since 2010 [1], a shortage of transplantable organs still remains. The dearth of available organs for transplantation is particularly acute among Latinx populations, an increasingly accepted and common term to represent Hispanic and Latino communities [2]. As of September 2019, over 23,000 Latinxs were candidates on the national transplant waitlist [3], of whom 88% were in need of a kidney transplant [4] because of higher rates of diabetes, chronic kidney disease, and end-stage renal disease compared with non-Hispanic whites. The need among older Latinxs is particularly acute. A large longitudinal study (n=453,162) of kidney transplant candidates showed that the proportion of Latinxs aged 45 years and older on the waitlist grew consistently from 56.4% during the 1995 to 1999 period to 69.8% in the 2010 to 2014 period [5]. Despite the need, only 1508 (14%) deceased organ donors were Latinxs in 2018 [6], although interest in and willingness to donate have been expressed within Latinx populations, including uninsured undocumented Latinx immigrants [7,8].

Since the 2007 revision of the Uniform Anatomical Gift Act making the act of registering as an posthumous organ donor legally binding [9], much like an advance directive or living will, it has become an even more critical first step to increasing the number of individuals converted to actual donors and the number of organs available for transplant. Currently, approximately half of the US population have designated themselves as posthumous organ donors through drivers' licenses, donor cards, or Web-based registries—an act referred to as donor designation [10]. Donor designation rates among Latinxs have remained much lower than the national average. For instance, a retrospective cohort study of deceased donation decisions found that 6.4% of Latinx participants were designated organ donors [11], and a larger study (n=2070) found that less than 40% of Latinxs had documented their intentions to donate posthumously [12]. Although Latinxs in the United States come from various countries of origin and have distinct immigration histories, the extant literature describes this population as a homogeneous unit. Commonly cited barriers to donation among this population include low levels of acculturation, religious beliefs, mistrust of the medical and organ distribution systems, and cultural taboos regarding discussions about death [13-15]. Encouraging donor designation among all ethnic groups is imperative to improve access to transplantation for all patients on the waitlist. However, increasing the number of Latinx donors is of particular import, given the improved graft outcomes associated with receipt of organs from this subgroup of the population [16].

Prior research has underscored the importance of culturally targeted approaches to addressing the specific health beliefs and transplantation-related needs among Latinx communities. For example, Hu et al's family-based intervention on diabetes self-management among Latinxs yielded successful outcomes, such as improved blood pressure, higher knowledge about diabetes, healthier behaviors, and lowered body mass index [17]. Gordon et al created and evaluated a culturally targeted website about living kidney donation and transplantation for Latinx patients; exposure to the website in combination with transplant education generated significantly higher transplant knowledge compared with transplant education alone [18]. Other research aimed at increasing organ donation among Latinx populations has suggested women as ideal champions for promoting donor designation, given their dominant position in health-related decision making and their increased willingness to donate and discuss donation with family members [14,19,20].

Promotores de Salud

Promotores de Salud are a network of Latinx lay health educators (or community health workers) who disseminate health information to their communities through interpersonal channels. Indeed, platicas, or small group discussions typically held in homes or local community centers about different health topics, are organized and guided by promotores. In the 1960s, Promotores de Salud emerged out of necessity in low-income neighborhoods with sizable Latinx immigrant populations, to overcome significant barriers to accessing and utilizing health care services [21]. Promotores are a cost-efficient and effective workforce that expands the reach of the health care delivery system. As such, promotores have typically become the first point of contact between Latinx community members and formal health institutions. Given their knowledge of and membership in local and health care communities, promotores serve as cultural intermediaries, social/emotional supports, peer advocates, connections to available services, and catalysts for community action [22,23].

Promotores de Salud have had demonstrable impacts on health knowledge and behaviors. For instance, they have increased Latinxs' knowledge of and participation in preventive behaviors for diabetes [24], cervical cancer [25,26], obesity [27,28], and cardiovascular disease [29]. Furthermore, a growing body of evidence suggests promotores can be effectively engaged in the research process and, once trained, can be successful partners in study implementation [30-35]. Promotores have been trained to provide culturally and linguistically appropriate information to Latinx Americans about a wide variety of health-related topics and to promote behaviors to improve health and prevent illness, disease, and disability [36,37]. As more than 80% of these community health workers are female [36] and older Latinxs are increasingly overrepresented on the national waitlist [5], we believed that this existing network of promotores could be leveraged to educate mature Latinas (aged 50 years and older) in their respective communities about organ donation. We also anticipated that mature Latinas would be influenced to designate themselves as posthumous organ donors and support donation

within their families, as women are the primary source of information about health and health care in Latinx families. In addition, women, in particular, express greater willingness to register as deceased organ donors [14,19] and are more likely to register, be converted to actual donors, and discuss donation with their families [14,19,20].

Web-Based eLearning Modules

Web-based electronic learning (eLearning) platforms (ie, learning management systems) are highly variable in their features and methods of instruction [38]. Such customizability allows developers to remain sensitive to the practical and cultural needs of the target audience and helps ensure ease of use and engagement with material. Interactive computer- and Web-based public health and patient education trainings [38,39] have successfully addressed the learning needs of older adults [40-42], underserved populations [18,43-48], and those with chronic communicable disease [49], especially when culturally informed [50]. Previous studies employing Web-based interventions with targeted features to effectively train *promotoras* have demonstrated that eLearning leads to wider reach [51] and impact in Latinx communities [52]. The Office of Minority Health (OMH) of the US Department of Health and Human Services' creation and promotion of a Web-based training for *promotoras* about best practices for promoting health in their communities underscores this fact [53].

Promotoras de Donación

We developed the *Promotoras de Donación* Web-based eLearning module to increase *promotoras'* knowledge of organ donation and transplantation and to improve their communication skills and confidence discussing and promoting organ donation with mature Latinas. The feminine form of the term *promotor* was chosen because most community health workers are female, as are our target audiences. The module is designed to increase rates of donor designation within Latinx communities across the United States and to ultimately reduce disparities in transplantation among Latinxs. This paper describes our development of the *Promotoras de Donación* eLearning module, including the formative work involved in the module's development and the module's theoretical underpinnings. We present the module design, structure, cultural foundations, the results of beta-testing, and the plan for modifying the module in response to the feedback received. The paper concludes with challenges to the development of the module; limitations; and plans to evaluate the module's direct, indirect, and sustained effects.

Methods

eLearning Module Development

Study Team

The research team comprised Temple University faculty and staff including the principal investigator with expertise in interpersonal health communication and organ donation and transplantation (HG), 2 coinvestigators—a medical anthropologist with expertise in developing culturally targeted interventions in organ transplantation and donation from Northwestern University (EJG) and an internationally recognized expert in deceased organ donation (LS)—a creative coordinator (GPA), a project coordinator (CA and JT), and an undergraduate research assistant (PJK). The Gift of Life Institute, a Philadelphia-based international training center with extensive experience designing and developing Web-based educational programs for donation professionals, aided in curriculum design and module interface development (TD and RN). Textbox 1 illustrates the steps in the module's development.

Textbox 1. Steps in module development, implementation, and evaluation.

Formative research (September 2016–August 2017)

1. Recruited 4 promotore organizations as community-based partners.

2. Conducted cognitive interviews to develop and refine focus group moderator's guides.

3. Conducted and analyzed 12 focus group discussions—8 with mature Latinas and 4 with promotoras (N=98).

Module development (September 2017–August 2018)

4. Theoretically grounded in Organ Donation Model and Vested Interest Theory.

5. Curriculum development—integrated theories, extant literature, and formative research findings.

6. Script writing—identified didactic and skills-based sections of the module and strategically placed embedded media.

7. Design features developed in consultation with Gift of Life Institute and ensured cultural and linguistic targeting

8. Filming preparation including hiring and training actors and coordinating with production team.

9. Filming, editing, and embedding content in learning management system.

Implementation (September 2018–December 2018)

10. Soft launch for testing with Temple University team.

11. Beta-test (N=10) for acceptability, cultural appropriateness, platform navigability.

12. Integration of beta-test results into final version of module.

13. Assessment of direct effects of module on Promotoras' (N=40) organ donation attitudes and knowledge, donor registration behaviors, and confidence discussing donation.

14. Assessment of indirect effects of module on mature Latinas' organ donation knowledge and attitudes, and donor registration behaviors.

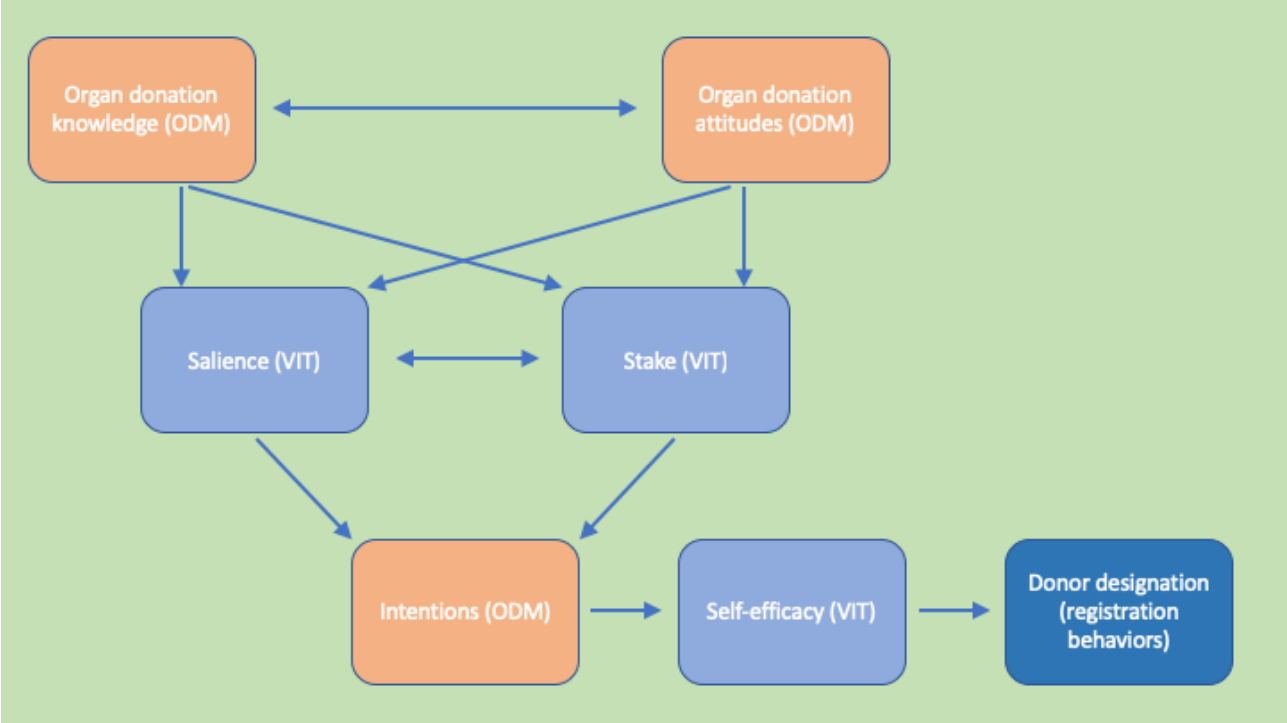
15. Assessment of sustained effects of module on Promotoras' organ donation attitudes and knowledge, donation behaviors, and confidence discussing donation

We partnered with 4 *promotoras* organizations representing 3 geographically diverse locales and heterogeneous Latinx populations: the Bexar County Community Health Collaborative (San Antonio, Texas), Enlace Chicago and Centro San Bonifacio (Chicago, Illinois), and Esperanza Health Center (Philadelphia, Pennsylvania). Each site was selected based on established relationships with local *promotoras* organizations, cost, and representation of Latinxs from different countries of origin. Leadership and *promotoras* from each organization met virtually with the Philadelphia-based research team on a quarterly basis for updates on study progress and discussion of challenges and next steps. Partners also provided feedback on study processes including recruitment plans and materials, the content of study instruments, Spanish translations, branding, module content, and beta-testing. *Promotoras* organizations guided the peripheral, evidential, linguistic, and sociocultural strategies employed to ensure the resulting module was culturally targeted and linguistically congruent [54].

Theoretical Foundation

Promotoras de Donación is theoretically rooted in the Organ Donation Model (ODM) and Vested Interest Theory (VIT). Figure 1 provides a pictorial representation of the module's theoretical framework. Proposed by Morgan and Miller [55,56], the ODM is grounded in the Theory of Reasoned Action (TRA) and highlights the importance of donation intentions on the act of becoming a designated organ donor. The model posits that attitudes toward and knowledge about organ donation influence designation intentions [55,56]. The model has found support in empirical tests among non-Hispanic whites and, in a slightly modified form, among African Americans [57-60]. In addition, constructs from both Azjen's Theory of Planned Behavior, a descendent of the TRA, and VIT have been useful in explaining differences in perceptions about living and nonliving organ donation among Hispanics [61].

Figure 1. Theoretical framework of *Promotoras de Donación*. Organ Donor Model (ODM) Vested Interest Theory (VIT).



VIT similarly delineates the relationships among and between constructs thought to predict donor designation [61]. VIT has been used to help explain the discrepancy between highly positive attitudes toward donation and low rates of donor designation. Specifically, VIT posits that the relationship between donation-related attitudes and low donor designation rates is moderated by perceptions of personal importance (ie, vested interest). VIT delineates 5 dimensions of *vested interest*: *stake* (perceived positive and negative consequences associated with a given issue), *salience* (vividness and accessibility of the issue), *self-efficacy* (perceived ability to overcome behavioral barriers), *certainty* (belief that the perceived consequences will be realized), and *immediacy* (belief in the imminent manifestation of those consequences) [62]. Scholars recommend the application of VIT to understand organ donation designation behaviors, focusing on the first 3 dimensions and indicating that self-efficacy in designating oneself as a donor proved a strong predictor of donor registration [63-65]. We conceptualize VIT constructs of stake, salience, and self-efficacy as representing awareness of the benefits of donation, recognition of the importance of and need for organ donors in the Latinx community, and confidence in making a decision about donation and donor designation.

These theoretical constructs are incorporated throughout the module. Short videos created by the Health Services and Resources Administration's Division of Transplantation as well as narratives (ie, anecdotal and expert testimony) and statistical messages (ie, data about the need for organ donors) are included to increase and enhance *promotoras'* knowledge and attitudes toward organ donation and donor designation. To increase self-efficacy, the module demonstrates the skills needed to introduce and manage *platicas* about organ donation through videos that model such discussions. VIT constructs are included in illustrations of the discrete skills needed to effectively present the need for Latinx donors (salience), articulation of the benefits of donation (stake), and presentation of instructions for donor registration (self-efficacy).

Module Structure and Curriculum

Promotoras de Donación was designed to be a highly engaging and interactive online learning experience (see Figure 2). Interactivity is an essential feature of learning in online environments and can support the pace at which the content is presented and enacted, development of associations between existing knowledge and acquisition of new information, reinforcement and refinement of newly acquired skills, guidance through the content learned, and enjoyment of the educational experience [66-68]. In addition, we accommodated multiple learning styles by incorporating a variety of pedagogic approaches into the module. For instance, written information and narration is accompanied by a public educational video promoting organ donation by depicting a canine pet following the deceased owner's transplanted heart to another person [69]; a video describing the organ donation process [70]; an animated video of how to register as an organ donor; a true/false activity; and video testimonials from a donor mother, a liver transplant recipient, and a transplant expert. See Multimedia Appendix 1 for snapshots of its general elements.

Figure 2. Screenshot of the login page of *Promotoras de Donación* to illustrate the branding standards.

Produced in Spanish (with English subtitles), the module was designed to achieve 3 specific learning objectives. After completing the module, we anticipated that learners would (1) demonstrate high levels of knowledge of organ donation and donor designation; (2) confidently engage members of their community in conversations about organ donation; and (3) effectively persuade individuals to designate themselves as posthumous organ donors via a donor card, drivers' license, or online registry. Guided by a female narrator *(narradora)*, the module originally comprised 6 sections (*secciones*): (1) introduction to organ donation and transplantation, (2) how to register as an organ and tissue donor (3) how to talk with others about organ and tissue donation, (4) a true/false activity, (5) how to promote donor registration, and (6) conclusion. Table 1 lists the subjects covered within each section and the embedded media before the final version of the module after beta-testing.

The curriculum consists of 2 components: a skills-based communication component and a didactic educational component. The communication skills needed to effectively discuss donation and share educational content were modeled in videos depicting a *promotora* leading a *platica* about organ donation and transplantation and promoting donor registration (see Figure 3). We employed and trained 7 *promotoras* from our local partnering organization to portray the roles of the *promotora* and *platica* attendees. In collaboration with the partnering organizations, scripts were developed to train the *promotora* actors and incorporate cultural concerns revealed in the formative focus group discussions throughout the module. The actors were trained weekly in October and November 2017, and the videos were filmed in December 2017. The videos and activities were edited, rendered, and embedded within a learning management system hosted by the Gift of Life Institute.

The skills-based component provides instruction on the communication skills needed to effectively engage participants during *platicas* about organ donation. To be effective communicators, *promotoras* must acquire the skills needed to discuss organ donation and promote donor designation and be motivated to use the skills [71,72]. The *platicas* depicted in the module are intended to model these skills and build communication self-efficacy or confidence in starting the discussion about donation, addressing concerns raised by attendees, and promoting the act of donor designation using persuasive but noncoercive language. In addition, the program trains *promotoras* to exercise communicative tasks such as broaching the topic of organ donation, providing basic information, highlighting the benefits of donation to society, emphasizing the need for Latinx donors, promoting and addressing concerns about donor designation, providing instruction on how to register as an organ donor, answering questions, and closing the discussion.

Table 1. *Promotoras de Donación* module content.

Section and content	Embedded media
Section 1: Introduction to Organ Donation and Transplantation	
Review of learning outcomes	Argentinian PSA[a] depicting an older man and his dog
Benefits of donation to donor, recipient, donor family, society	HRSA[b] Video "Organ Donation and Transplantation: How Does it Work?"
Information on the need for donors, the national transplant waiting list, the donation process	—[c]
Section 2: How to Register as an Organ and Tissue Donor	
Information on how to register as an organ donor and the importance of family discussion about donor designation	Testimonial: Latinx organ recipient
Section 3: How to Talk to Others About Organ and Tissue Donation	
How to prepare for a platica	Platica dramatization
Communication skills: assess existing knowledge, ask open-ended and probing questions, provide factual information, correct myths and misinformation, and invite questions	Testimonial: transplant physician
Addressing myths: medical negligence, irreversibility of brain death, moral failure of organ recipient, undocumented immigrant eligibility, religious objections, black market organ trafficking	Testimonial: Latina mother of a deceased donor
How to converse with the family about end of life decisions	—
Section 4: A True/False activity	
Statements concern: medical negligence, prohibitive religious beliefs, prohibitive age and illness status, donation costs, infringed funeral wishes, inelligibility of the undocumented person, precedence of advanced directives	True/False activity
Section 5: How to Promote Donor Registration	
Communication skills: gauging support, emphasize importance, correct myths and misinformation, support individual choice, remain nonjudgmental, provide instruction on how to register	Second part of platica dramatization
Methods of donor designation: DMV[d], online, a donor card	—
Addressing common reasons people do not designate	—
How to close the platica	—
Section 6: Conclusion	
Reminder to revisit *Promotoras de Donación* as needed	—

[a]PSA: public service announcement.

[b]HRSA: Health Resources and Services Administration.

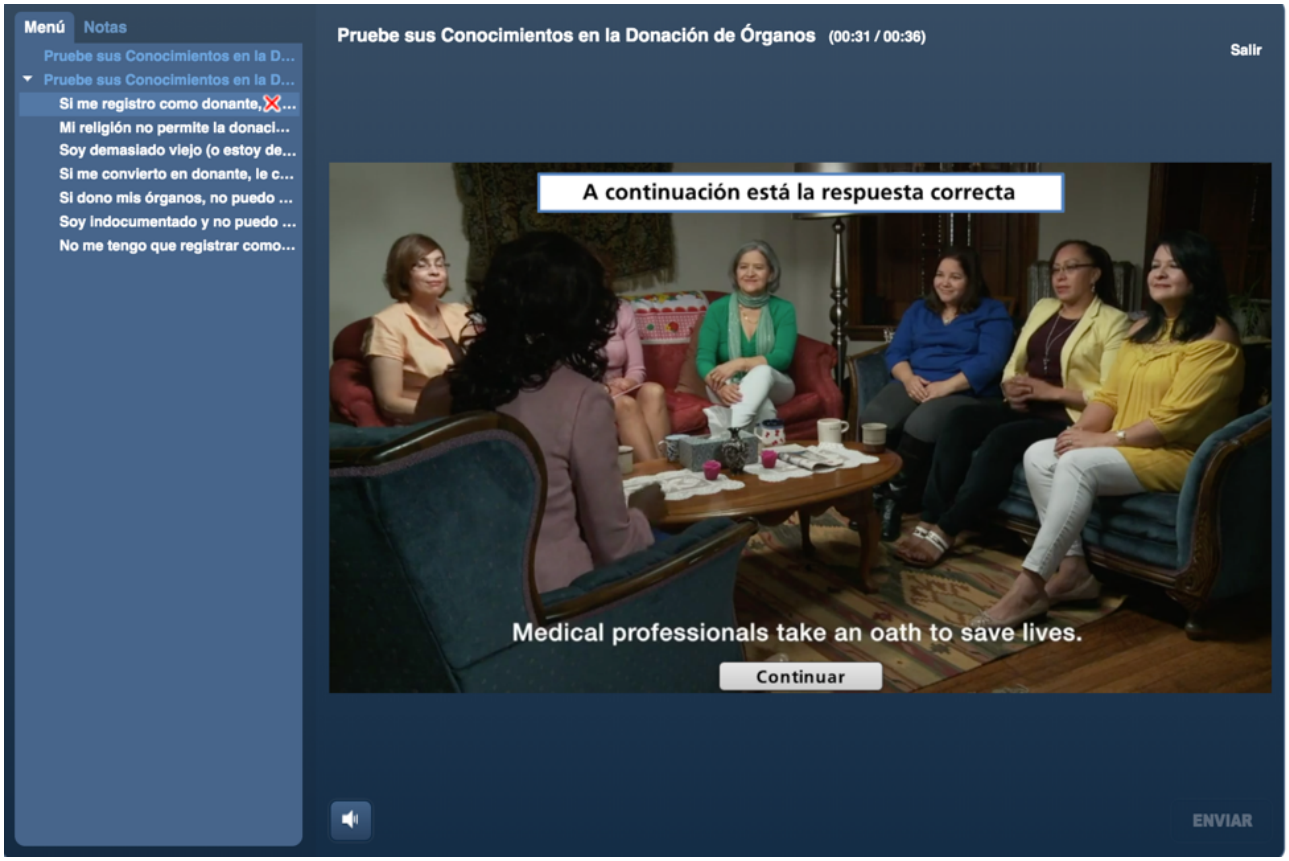
[c]Not applicable.

[d]DMV: Department of Motor Vehicles.

The didactic component provides basic information about organ donation and transplantation and the need for donors in the Latinx community. It also addresses concerns about organ donation and donor designation commonly shared by the target population, as revealed through the extant literature and a series of 12 focus groups we conducted with *promotoras* (n=37) and older Latinas (n=61) to inform the module's development [73]. Focus group participants reported myths and misinformation about organ donation, medical mistrust, and family aversion of discussing sensitive topics as some of the major barriers to organ donation and donor designation. Participants also acknowledged the need for transplantable organs and increased education about the topic within Latinx communities.

The didactic component featured several topics during the dramatized *platicas*. *Platica* attendees asked about stories they heard about the black market, beliefs about health care providers doing less to save the lives of designated organ donors, worries about the incompatibility of organ donation and religious beliefs, the impact of being undocumented immigrants on the ability to receive a transplant, and how to handle discussions about organ donation with family members. The module depicts the *promotora* skillfully and thoughtfully responding to the attendees' questions and fears about organ donation in Latinx communities and drew on their shared experiences and cultural concerns to deliver information about organ donation and to promote donor designation. The true/false activity was designed to test retention of and reinforce the information provided throughout the module.

Figure 3. Screenshot of a platica dramatization from the *Promotoras de Donación*.



A primary goal of module development was ensuring its cultural appropriateness and receptivity by the target population. Throughout the scriptwriting and postproduction process, Spanish-speaking Latinx consultants and our Latinx community partners offered guidance to ensure that the module and its content were realistic, culturally targeted, and linguistically congruent [46]. Our collaborators helped to refine the dialog and educational material by providing insight on focus group data reflecting cultural beliefs and myths about donation. *Promotora* actors represented different Latin American countries and accents to model realistic donation discussions that reflect the variety of Latinx nationalities represented in the United States. The module's script also incorporated specific statistics about the need for donors in the Latinx community. Table 2 details the specific design and content features that were included to ensure the module's cultural and linguistic appropriateness [55].

A certificate of completion is provided upon module conclusion. The *Promotoras de Donación* module is accompanied by an Educator's Manual containing a list of potential discussion topics, commonly asked questions about donation and their answers, and sample language to sensitively correct myths about donation.

Table 2. Strategies for achieving cultural and linguistic appropriateness.

Strategy	Design feature	Content feature
Peripheral	Branding and color schema	Use of Latina narrator (narradora)
Evidential	—[a]	Videos and testimonials from Latina organ recipient, donor family, and transplant physician
Linguistic	—	Spanish with English subtitles; available Spanish on screen text
Constituent involving	Feedback from promotoras and leadership from promotora organizations	Integrated feedback from promotoras and leadership from promotora organizations, promotoras as module actors
Sociocultural	—	Module addresses commonly cited myths about organ donation found in literature and focus group data

[a]Not applicable.

Beta-Testing

Procedure

The original version of the module was timed at 1 hour, 14 min, and 29 seconds. The module underwent beta-testing from September 17, 2018, to October 12, 2018. Beta-testing involved full navigation of the module and completion of a postmodule questionnaire measuring user experience and soliciting feedback for improvement. A total of 10 people participated in beta-testing including 3 members of the leadership teams at partnering *promotoras* organizations, 6 *promotoras*, and 1 partnering organization administrator. *Promotoras* participating in beta-testing were compensated with a US $25.00 gift card.

Survey

The beta-test questionnaire comprised 40 questions evaluating perceptions of the module; cultural sensitivity; platform navigability; and sociodemographics including job role, length in job role, and preferred language spoken at home. See Multimedia Appendix 2 for the beta-test questionnaire.

Perceptions of the module were assessed through 13 5-point Likert questions (1—*strongly disagree*; 5—*strongly agree*) ascertaining organization, length, realism, and interest in the material covered. Cultural competency was assessed through 5 4-point Likert-type questions (1—*strongly disagree*; 4—*strongly agree*) adapted from the Cultural Sensitivity Assessment Tool [74]. Participants were asked to assess website and module navigability using 5 4-point Likert-type questions (1—*strongly disagree*; 5—*strongly agree*) adapted from the Informational Fit-to-Task Items [75].

Participants were then asked to identify which segments (ie, true/false activity and testimonials) could be shortened or removed through 8 3-point Likert-type questions (1—*definitely remove*; 3—*definitely do not remove*). A single 5-point Likert-type item measured respondents' rating of the overall quality of the module (1—*poor*; 5—excellent). The questionnaire ended with 5 open-ended questions that asked respondents to indicate suggestions for improving the training program, features that were missing or did not feel right, the most important message for people unfamiliar with the topic of organ donation and transplantation, and the most and least liked aspects.

Results

Quantitative Results

Most beta-testers spoke Spanish and English equally at home (7/10) and identified as *promotoras* (6/10); the remaining 4 were leadership from the 4 partnering organizations. Most respondents somewhat or strongly agreed that the information presented was novel (7/10), felt the module activities helped them learn (9/10), enjoyed module activities (9/10), felt ready to teach others about organ donation and transplantation after viewing the module (8/10), would recommend this training to other *promotoras* (9/10), and felt the module was too long (6/10). Most respondents (9/10) agreed or strongly agreed that the module was free from stereotypes about the Latinx community, that the people in their community would believe the messages in the module were from credible sources, and that the messages in the module address commonly shared organ donation and organ transplantation myths in the Latinx community. Regarding website navigability, all respondents disagreed with the statement, "the website was too complex"; all respondents agreed or strongly agreed that the website was easy to use and that they liked the look and feel of the website. Most respondents (9/10) disagreed that the words, phrases, and expressions used in the module were too technical for the average *promotora*, and most respondents (8/10) agreed or strongly agreed that *promotoras* would find the website easy to use.

Half of the respondents (5/10) suggested removing the true/false activity, and 40% (4/10) of the respondents suggested removing the interview with a transplant expert. Most (8/10) respondents recommended definitely not removing the segments describing personal experiences with donation (mother of an organ donor) and transplantation (liver transplant recipient).

Qualitative Results

In addition to scale items, responses to open-ended questionnaire questions provided rich, contextualizing data suggesting improvements to the original module. Table 3 presents the qualitative questions and a selection of corresponding responses.

We gleaned several lessons from themes emerging from the open-ended responses. First, although beta-testers stated the module was informative and interactive, most found the module to be too lengthy and offered suggestions on content that could be shortened or edited out. Moreover, they noted that the most important messages of the module included life-saving opportunities of organ donation and the need for Latinx donors; they also liked the testimonials included in the module.

Table 3. Selected responses to open-ended questions.

Open-ended question	Selected responses
Please use the space below to write down any suggestions on ways to improve the training program.	• "This was very well done. I learned a lot and registered as an organ donor right at the beginning after the video with the dog." • "I wouldn't entirely remove the videos with the doctor and the mother who lost her son, but would recommend cutting them down. I'd also cut down some of the conversation led by the promotora. That section was definitely too long." • "The module is very well done, with interesting ways to present the topics and with several personal stories. I did not feel that it was boring. To improve it, I wonder if there is any way to connect the promoters [sic] with a local organization if they want to learn more or become more involved in this area."
What, if anything, was missing or did not feel right for you in this module?	• "I can't think of anything really. There was just the length that was way too long for me." • "I was not sure if the promotora was saying "traNsplante" or "trasplante". The correct pronunciation would be trasplante for a transplant."
What do you think is the most important message in this video for people who are unfamiliar with the topic of organ donation and transplantation?	• "One person can make a difference." • "That organ donation/transplantation saves lives and is definitely something to consider and talk to others about." • "The video explains the importance of organ donation in the Latino community and the need that exists in the community of more Latino organ donors." • "The myths. The amount of persons a donor can help and tissue, skin, etc. donations"
What did you like most about the module?	• "I like that the module is interactive and presents the information in several ways." • "The information that older people [can be] included as donors." • "The real people interviews." • "The participation of the people in the talk and the confidence that the promoter gave." • "Informative in a relaxed, friendly environment"
What did you like least about the module?	• "A lot of explanations from the host... some of the content was then repeated in the videos (e.g., about how to register as an organ donor)." • "It needs to be cleaned up a bit to be a bit shorter."

Discussion

Principal Findings

Promotoras de Donación is the first online training program about organ donation designed specifically for promotoras. We employed a rigorous process to develop this eLearning module. Specifically, following well-established guidelines, we incorporated findings from the extant literature research on organ donation beliefs among Latinx Americans, relied on established theories for guidance, and engaged promotoras and leadership from promotora organizations for advice and feedback throughout the process. As no research had yet examined the knowledge of and attitudes toward organ donation and transplantation or the act of donor designation among either promotoras or mature Latinas—our target audiences—we conducted a series of focus group interviews to elicit this information [74]. The findings of the group interviews were integrated in the design and content of the module. The result was the creation of an empirically and theoretically based, culturally and linguistically appropriate eLearning module designed through a collaborative partnership with promotoras and promotora organizations.

Beta-testing offered valuable feedback about user experience, acceptability, and content. To prepare the module for implementation and evaluation, we made several modifications. For example, we modified the module by increasing the font size for the true/false activity, enabling sensible advancement to subsequent sections for ease of module progression, including removing numbered sections on the menu bar, adding English subtitles, and disabling user manipulation of the progress bar. To reduce module length, we edited testimonials from the mother who had donated her deceased son's organs and the transplant expert. We moved the animation depicting 3 ways to donate and the true/false activity to appear on the homepage after module completion as supplemental, rather than required material. After revisions, the *Promotoras de Donación* module lasts 1 hour, 3 min, and 38 seconds, with 7 min and 49 seconds of supplemental content. Table 4 provides a comparison of the original and final versions of the module.

Table 4. Design elements in original and final versions of the module.

Design element	Original version	Final version
Subtitling	Absent	• Added in English with transcript of script available at all times in a togglable tab
Advancement through the module	Unclear and confusing numbering	• Logical numbering of each section • Explicit instructions to progress through module
Progress bar	User able to manipulate	• Disabled ability to manipulate
Older man and the dog video	Section 1	• No change
HRSA[a] video: "Organ Donation and Transplantation: How Does it Work?"	Section 1	• No change
Testimonial: Latinx organ recipient	Section 2	• No change
Platica dramatization	Sections 3 and 5	• Appear closer together in sections three and four
Testimonial: transplant physician	Section 3	• Retained in section three but tailored testimonial to address concerns of a black market to reduce length
Testimonial: Latina mother of a deceased donor	Section 3	• Retained in section three but reduced in length
True/False activity	Section 4	• Removed from module and included as a supplemental activity • Font sized increased
Three ways to donate animation	Embedded within the module	• Removed from module and included as a supplemental activity

[a]HRSA: Health Resources and Services Administration.

Next Steps

The module is currently undergoing evaluation of its direct, indirect, and sustained effects. In the first phase of evaluation, we will recruit 40 *promotoras* (10 per partnering site) to complete the *Promotoras de Donación* module and associated pre- and postsurveys. The presurvey will be administered before participants are given access to the module, and the postsurvey will be administered upon module completion; both surveys have been translated into Spanish and will be administered online via Qualtrics. The surveys assess changes in organ donation knowledge and attitudes and confidence communicating about and promoting the act of donor designation.

After completing phase 1, the same *promotoras* will be invited to enroll in the second phase of the evaluation, conducted to test the indirect effects of the module on mature Latinas' organ donation knowledge, attitudes, and intentions to register as posthumous organ donors. Specifically, trained *promotoras* will lead 2 *platicas* in their respective communities with up to 8 mature Latinas in each session. Before leading the *platicas*, participating *promotoras* will be apprised of the study protocols and trained in human subject's protections. *Promotoras* will administer anonymous paper-pencil surveys to mature Latinas before beginning the discussion and before attendees leave the *platica*.

To measure the module's sustained impact on *promotoras'* organ donation knowledge and attitudes as well as communication confidence, participating *promotoras* will complete self-administered surveys before leading their first *platica* and after completing both. If the intervention demonstrates significantly increased rates of donor designation, our Web-based training will be disseminated widely to *promotoras* organizations across the United States and a petition will be made to the HHS OMH to include the module in its existing eLearning program for *Promotores de Salud* [53].

Challenges

Several challenges arose in the development of *Promotoras de Donación* module. Reliance on multiple sources (ie, focus groups, partner organizations, and beta-testers) for input on module content and format increased module production and revision time frames. Although the feedback was critical to ensure the cultural appropriateness and relevance of the module for the target population, we underestimated the time needed to solicit, review, and incorporate feedback from every collaborating partner. Turnover in staff also contributed to the delays. Our original creative coordinator, who was a fluent Spanish speaker with advanced training in research, resigned necessitating the search for Spanish-speaking consultants in each state to facilitate the focus groups and analyze the resulting data. We also lost 2 study coordinators during critical periods of module development—the first was during finalization of the original version and the second was just before final modifications were made. Other researchers developing similar community-based eLearning interventions would do well to anticipate longer production timetables by adding 2 to 3 months before the launch of the intervention.

The decision to incorporate multiple sources of feedback on the module's content also led to a longer overall running time than was originally planned. Although the final version of *Promotoras de Donación* was shortened by making original components supplemental, the module still lasts over 63 min in its final form with close to 8 min of supplemental material. This is considerably longer than initial plans of an approximately 20 to 30 min eLearning module. Even though we have yet to see if the increased length of the module will have any demonstrable impact on participant's experience and satisfaction, the substantial increase in content required additional time for editing, additional hosting space to accommodate the size of media files, and troubleshooting with the Gift of Life Institute's technological capabilities.

Conclusions

Promotoras de Donación is a novel, culturally targeted intervention firmly grounded in theory, extant literature, and formative research. In response to a lack of knowledge of organ donation and low willingness to register as a posthumous organ donor among Latinxs, the module aims to engender more favorable attitudes toward organ donation, improve knowledge levels, and build confidence discussing donor designation among promotoras as means of increasing registration rates within Latinx communities. The results of our beta-testing demonstrated strong user acceptability of the platform and the content. Furthermore, the module may be dispersed to Latinx communities across the nation at low cost and has considerable potential to increase the number of Latinx registered as deceased organ donors, thus helping to reduce the ongoing national shortage of transplantable organs.

Acknowledgments

This study was supported by the Health Resources and Services Administration's Division of Transplantation (HRSA/DoT R39OT29878, PI: Traino). The content is solely the responsibility of the authors and does not necessarily represent the official views of HRSA/DoT.

Authors' Contributions

All authors participated in research design, writing of the paper, performance of the research. HG, EG, PJK, and JT participated in data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the Promotoras de Donación eLearning module.
[PDF File (Adobe PDF File), 19215 KB - jmir_v22i1e15793_app1.pdf]

Multimedia Appendix 2
Post Module Questionnaire.
[PDF File (Adobe PDF File), 64 KB - jmir_v22i1e15793_app2.pdf]

References

1. OPTN: Organ Procurement and Transplantation. Donors Recovered in the US by Donor Type: Donors Recovered: January 1, 1988 - August 31, 2019, Based on OPTN data as of September 23, 2019 URL: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/# [accessed 2019-09-23]
2. Salinas C, Lozano A. Mapping and recontextualizing the evolution of the term Latinx: an environmental scanning in higher education. J Latinos Educ 2017;18(4):302-315 [FREE Full text] [doi: 10.1080/15348431.2017.1390464]
3. OPTN: Organ Procurement and Transplantation. Waitlist: Ethnicity by Gender: Current US Waiting list, Based on OPTN data as of September 23, 2019 URL: https://optn.transplant.hrsa.gov/data/view-data-reports/build-advanced/ [accessed 2019-09-23]
4. OPTN: Organ Procurement and Transplantation. Organ by Ethnicity: Current US Waiting List, Based on OPTN data as of September 23, 2019 URL: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/# [accessed 2019-09-23]
5. Purnell TS, Luo X, Cooper LA, Massie AB, Kucirka LM, Henderson ML, et al. Association of race and ethnicity with live donor kidney transplantation in the United States from 1995 to 2014. J Am Med Assoc 2018 Jan 2;319(1):49-61 [FREE Full text] [doi: 10.1001/jama.2017.19152] [Medline: 29297077]
6. OPTN: Organ Procurement and Transplantation. Deceased Donors Recovered in the US by Donor Ethnicity: Donors Recovered: January 1, 1988 - August 31, 2019, Based on OPTN data as of September 23, 2019 URL: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/# [accessed 2019-09-23]

7. Purnell TS, Powe NR, Troll MU, Wang N, LaVeist TA, Boulware LE. Donor designation: racial and ethnic differences in US nondesignators' preferred methods for disclosing intent to donate organs. Transpl Int 2011 Oct;24(10):999-1007 [FREE Full text] [doi: 10.1111/j.1432-2277.2011.01301.x] [Medline: 21777299]
8. Baru JS, Lucas BR, Martinez C, Brauner D. Organ donation among undocumented hispanic immigrants: an assessment of knowledge and attitudes. J Clin Ethics 2013;24(4):364-372. [Medline: 24597424]
9. Verheijde JL, Rady MY, McGregor JL. The United States Revised Uniform Anatomical Gift Act (2006): new challenges to balancing patient rights and physician responsibilities. Philos Ethics Humanit Med 2007 Sep 12;2:19 [FREE Full text] [doi: 10.1186/1747-5341-2-19] [Medline: 17850664]
10. Donate Life America. 2017. Donate Life America: 2017 Annual Update URL: https://www.donatelife.net/wp-content/uploads/2016/06/2017_AnnualUpdate_singlepages_small.pdf [accessed 2019-10-08]
11. Traino HM, Siminoff LA. Attitudes and acceptance of First Person Authorization: a national comparison of donor and nondonor families. J Trauma Acute Care Surg 2013 Jan;74(1):294-300 [FREE Full text] [doi: 10.1097/TA.0b013e318270dafc] [Medline: 23147186]
12. Sehgal NK, Scallan C, Sullivan C, Cedeño M, Pencak J, Kirkland J, et al. The relationship between verified organ donor designation and patient demographic and medical characteristics. Am J Transplant 2016 Apr;16(4):1294-1297 [FREE Full text] [doi: 10.1111/ajt.13608] [Medline: 26603147]
13. Salim A, Schulman D, Ley EJ, Berry C, Navarro S, Chan LS. Contributing factors for the willingness to donate organs in the Hispanic American population. Arch Surg 2010 Jul;145(7):684-689 [FREE Full text] [doi: 10.1001/archsurg.2010.110] [Medline: 20644132]
14. Breitkopf CR. Attitudes, beliefs and behaviors surrounding organ donation among Hispanic women. Curr Opin Organ Transplant 2009 Apr;14(2):191-195 [FREE Full text] [doi: 10.1097/MOT] [Medline: 19469039]
15. Frates J, Bohrer G. Hispanic perceptions of organ donation. Prog Transplant 2002 Sep;12(3):169-175. [Medline: 12371041]
16. Kwan JM, Hajjiri Z, Chen YF, Metwally A, Perkins DL, Finn PW. Donor and recipient ethnicity impacts renal graft adverse outcomes. J Racial Ethn Health Disparities 2018 Oct;5(5):1003-1013. [doi: 10.1007/s40615-017-0447-9] [Medline: 29270842]
17. Hu J, Wallace DC, McCoy TP, Amirehsani KA. A family-based diabetes intervention for Hispanic adults and their family members. Diabetes Educ 2014;40(1):48-59 [FREE Full text] [doi: 10.1177/0145721713512682] [Medline: 24248832]
18. Gordon EJ, Feinglass J, Carney P, Ramirez D, Olivero M, O'Connor K, et al. An interactive, bilingual, culturally targeted website about living kidney donation and transplantation for hispanics: development and formative evaluation. JMIR Res Protoc 2015 Apr 20;4(2):e42 [FREE Full text] [doi: 10.2196/resprot.3838] [Medline: 25896143]
19. Alvaro EM, Jones SP, Robles AS, Siegel JT. Predictors of organ donation behavior among Hispanic Americans. Prog Transplant 2005 Jun;15(2):149-156. [doi: 10.7182/prtr.15.2.548p24717r135842] [Medline: 16013463]
20. Organ Donor. 2013 Sep. 2012 National Survey of Organ Donation Attitudes and Behaviors URL: https://www.organdonor.gov/sites/default/files/about-dot/files/nationalsurveyorgandonation.pdf [accessed 2019-10-08]
21. U.S. Department of Health and Human Services, Indian Health Services. 2019 Jan 01. Community Health Representatives (CHR) Fact Sheet URL: https://www.ihs.gov/sites/chr/themes/responsive2017/display_objects/documents/chrfactsheet2018.pdf [accessed 2019-10-08]
22. Ingram M, Reinschmidt KM, Schachter KA, Davidson CL, Sabo SJ, de Zapien JG, et al. Establishing a professional profile of community health workers: results from a national study of roles, activities and training. J Community Health 2012 Apr;37(2):529-537 [FREE Full text] [doi: 10.1007/s10900-011-9475-2] [Medline: 21964912]
23. Rosenthal EL, Wiggins N, Ingram M, Mayfield-Johnson S, de Zapien JG. Community health workers then and now: an overview of national studies aimed at defining the field. J Ambul Care Manage 2011;34(3):247-259. [doi: 10.1097/JAC.0b013e31821c64d7] [Medline: 21673523]
24. Lujan J, Ostwald SK, Ortiz M. Promotora diabetes intervention for Mexican Americans. Diabetes Educ 2007;33(4):660-670. [doi: 10.1177/0145721707304080] [Medline: 17684167]
25. Byrd TL, Wilson KM, Smith JL, Coronado G, Vernon SW, Fernandez-Esquer ME, et al. AMIGAS: a multicity, multicomponent cervical cancer prevention trial among Mexican American women. Cancer 2013 Apr 1;119(7):1365-1372 [FREE Full text] [doi: 10.1002/cncr.27926] [Medline: 23280399]
26. Fernández ME, Gonzales A, Tortolero-Luna G, Williams J, Saavedra-Embesi M, Chan W, et al. Effectiveness of Cultivando la Salud: a breast and cervical cancer screening promotion program for low-income Hispanic women. Am J Public Health 2009 May;99(5):936-943. [doi: 10.2105/AJPH.2008.136713] [Medline: 19299678]
27. Sosa E, Biediger-friedman L, Yin Z. Lessons learned from training of promotores de salud for obesity and diabetes prevention. J Health Dispar Res Pract 2013;6(1):1-13 [FREE Full text]
28. Staten LK, Cutshaw CA, Davidson C, Reinschmidt K, Stewart R, Roe DJ. Effectiveness of the Pasos Adelante chronic disease prevention and control program in a US-Mexico border community, 2005-2008. Prev Chronic Dis 2012;9:E08 [FREE Full text] [doi: 10.5888/pcd9.100301] [Medline: 22172175]
29. Balcázar H, Alvarado M, Cantu F, Pedregon V, Fulwood R. A promotora de salud model for addressing cardiovascular disease risk factors in the US-Mexico border region. Prev Chronic Dis 2009 Jan;6(1):A02 [FREE Full text] [Medline: 19080008]

30. Marsh T, Derose KP, Rios M, Cohen D. Promotoras as data collectors in a large study of physical activity in parks. Health Promot Pract 2015 May;16(3):354-361 [FREE Full text] [doi: 10.1177/1524839914563076] [Medline: 25649234]
31. Johnson CM, Sharkey JR, Dean WR, St John JA, Castillo M. Promotoras as research partners to engage health disparity communities. J Acad Nutr Diet 2013 May;113(5):638-642 [FREE Full text] [doi: 10.1016/j.jand.2012.11.014] [Medline: 23375463]
32. St John JA, Huber JC, Sharkey JR, Dean WR. Using direct observations on multiple occasions to measure household food availability among low-income Mexicano residents in Texas colonias. BMC Public Health 2010;12(445). [doi: 10.1007/s10935-013-0296-1] [Medline: 23404423]
33. O'Hegarty M, Pederson LL, Thorne SL, Caraballo RS, Evans B, Athey L, et al. Customizing survey instruments and data collection to reach Hispanic/Latino adults in border communities in Texas. Am J Public Health 2010 Apr 1;100(Suppl 1):S159-S164. [doi: 10.2105/AJPH.2009.167338] [Medline: 20147687]
34. Rhodes SD, Foley KL, Zometa CS, Bloom FR. Lay health advisor interventions among Hispanics/Latinos: a qualitative systematic review. Am J Prev Med 2007 Nov;33(5):418-427. [doi: 10.1016/j.amepre.2007.07.023] [Medline: 17950408]
35. Hanbury MM, Sadeghi B, Tseregounis IE, Gomez-Camacho R, Manzo RD, Rangel MI, et al. A web-based application to improve data collection in an interventional study targeting childhood obesity: pre-post analysis. J Med Internet Res 2019 Jan 16;21(1):e10861 [FREE Full text] [doi: 10.2196/10861] [Medline: 30664465]
36. Bureau of Health Workforce. 2007 Mar. Community Health Workers National Workforce Study URL: https://bhw.hrsa.gov/sites/default/files/bhw/nchwa/projections/communityhealthworkforce.pdf [accessed 2019-10-08]
37. Gonzalez M, Sanders-Jackson A, Wright T. Web-based health information technology: access among Latinos varies by subgroup affiliation. J Med Internet Res 2019 Apr 16;21(4):e10389 [FREE Full text] [doi: 10.2196/10389] [Medline: 30990462]
38. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ 2010 Aug;44(8):765-774. [doi: 10.1111/j.1365-2923.2010.03723.x] [Medline: 20633216]
39. Ryhänen AM, Siekkinen M, Rankinen S, Korvenranta H, Leino-Kilpi H. The effects of Internet or interactive computer-based patient education in the field of breast cancer: a systematic literature review. Patient Educ Couns 2010 Apr;79(1):5-13. [doi: 10.1016/j.pec.2009.08.005] [Medline: 19744817]
40. Neafsey PJ, Strickler Z, Shellman J, Chartier V. An interactive technology approach to educate older adults about drug interactions arising from over-the-counter self-medication practices. Public Health Nurs 2002;19(4):255-262. [doi: 10.1046/j.1525-1446.2002.19405.x] [Medline: 12071899]
41. Fox MP. A systematic review of the literature reporting on studies that examined the impact of interactive, computer-based patient education programs. Patient Educ Couns 2009 Oct;77(1):6-13. [doi: 10.1016/j.pec.2009.02.011] [Medline: 19345550]
42. Bond GE, Burr R, Wolf FM, Price M, McCurry SM, Teri L. The effects of a web-based intervention on the physical outcomes associated with diabetes among adults age 60 and older: a randomized trial. Diabetes Technol Ther 2007 Feb;9(1):52-59 [FREE Full text] [doi: 10.1089/dia.2006.0057] [Medline: 17316098]
43. Bell M, MacDougall K. Adapting online learning for Canada's Northern public health workforce. Int J Circumpolar Health 2013;72 [FREE Full text] [doi: 10.3402/ijch.v72i0.21345] [Medline: 23971012]
44. Choi J, Bakken S. Web-based education for low-literate parents in Neonatal Intensive Care Unit: development of a website and heuristic evaluation and usability testing. Int J Med Inform 2010 Aug;79(8):565-575 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.05.001] [Medline: 20617546]
45. Andersen S, Andersen P, Youngblood N. Multimedia computerized smoking awareness education for low-literacy Hispanics. Comput Inform Nurs 2011 Feb;29(2 Suppl):TC35-TC42. [doi: 10.1097/NCN.0b013e318215543f] [Medline: 21372644]
46. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
47. Gordon EJ, Feinglass J, Carney P, Vera K, Olivero M, Black A, et al. A culturally targeted website for Hispanics/Latinos about living kidney donation and transplantation. Transplantation 2016 May;100(5):1149-1160. [doi: 10.1097/TP.0000000000000932] [Medline: 26444846]
48. Gordon EJ, Feinglass J, Carney P, Vera K, Olivero M, Black A, et al. A website intervention to increase knowledge about living kidney donation and transplantation among Hispanic/Latino dialysis patients. Prog Transplant 2016 Mar;26(1):82-91. [doi: 10.1177/1526924816632124] [Medline: 27136254]
49. Kurth AE, Chhun N, Cleland CM, Crespo-Fierro M, Parés-Avila JA, Lizcano JA, et al. Linguistic and cultural adaptation of a computer-based counseling program (CARE+Spanish) to support HIV treatment adherence and risk reduction for people living with HIV/ADIS: A randomized controlled trial. J Med Internet Res 2016 Jul 13;18(7):e195 [FREE Full text] [doi: 10.2196/jmir.5830] [Medline: 27417531]
50. 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 [FREE Full text] [doi: 10.2196/jmir.2095] [Medline: 22652979]
51. 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]

52. Balcázar H, Alvarado M, Hollen ML, Gonzalez-Cruz Y, Pedregón V. Evaluation of Salud Para Su Corazón (Health for your Heart) -- National Council of La Raza Promotora Outreach Program. Prev Chronic Dis 2005 Jul;2(3):A09 [FREE Full text] [Medline: 15963311]
53. Think Cultural Health. US Department of Health and Human Services Office of Minority Health URL: https://www.thinkculturalhealth.hhs.gov/education/promotores [accessed 2019-10-08]
54. Kreuter MW, Lukwago SN, Bucholtz RD, Clark EM, Sanders-Thompson V. Achieving cultural appropriateness in health promotion programs: targeted and tailored approaches. Health Educ Behav 2003 Apr;30(2):133-146. [doi: 10.1177/1090198102251021] [Medline: 12693519]
55. Morgan S, Miller J. Communicating about gifts of life: the effect of knowledge, attitudes, and altruism on behavior and behavioral intentions regarding organ donation. J Appl Commun Res 2002;30(2):163-178 [FREE Full text] [doi: 10.1080/00909880216580]
56. Morgan SE, Miller JK. Beyond the organ donor card: the effect of knowledge, attitudes, and values on willingness to communicate about organ donation to family members. Health Commun 2002;14(1):121-134. [doi: 10.1207/S15327027HC1401_6] [Medline: 11853207]
57. Morgan SE, Miller JK, Arasaratnam LA. Similarities and differences between African Americans' and European Americans' attitudes, knowledge, and willingness to communicate about organ donation. J Appl Social Pyschol 2003;33(4):693-715 [FREE Full text] [doi: 10.1111/j.1559-1816.2003.tb01920.x]
58. Morgan SE. The power of talk: African Americans' communication with family members about organ donation and its impact on the willingness to donate organs. J Soc Person Relat 2016;21(1):112-124 [FREE Full text] [doi: 10.1177/0265407504039845]
59. Morgan SE. Many facets of reluctance: African Americans and the decision (not) to donate organs. J Natl Med Assoc 2006 May;98(5):695-703. [Medline: 16749644]
60. Robinson DH, Perryman JP, Thompson NJ, Amaral S, Arriola KR. Testing the utility of a modified organ donation model among African American adults. J Behav Med 2012 Jun;35(3):364-374 [FREE Full text] [doi: 10.1007/s10865-011-9363-3] [Medline: 21698439]
61. Sivacek J, Crano WD. Vested interest as a moderator of attitude–behavior consistency. J Pers Soc Psychol 1982 Aug;43(2):210-221. [doi: 10.1037/0022-3514.43.2.210]
62. Crano WD, Prislin R. Components of vested interest and attitude-behavior consistency. Basic Appl Soc Psych 2011 Jun 22;17(1-2):1-21 [FREE Full text] [doi: 10.1080/01973533.1995.9646129]
63. Anker AE, Feeley TH. Are nonparticipants in prosocial behavior merely innocent bystanders? Health Commun 2011 Jan;26(1):13-24. [doi: 10.1080/10410236.2011.527618] [Medline: 21308580]
64. Anker A, Feeley T, Kim H. Examining the attitude-behavior relationship in prosocial donation domains. J Appl Soc Psychol 2010;40(6):1293-1324 [FREE Full text] [doi: 10.1111/j.1559-1816.2010.00619.x]
65. Quick BL, Anker AE, Feeley TH, Morgan SE. An examination of three theoretical models to explain the organ donation attitude--registration discrepancy among mature adults. Health Commun 2016;31(3):265-274. [doi: 10.1080/10410236.2014.947468] [Medline: 26305921]
66. Holmberg B. Theory And Practice Of Distance Education. London: Routledge; 1989.
67. Domagk S, Schwartz RN, Plass JL. Interactivity in multimedia learning: an integrated model. Comput Human Behav 2010 Sep 01;26(5):1024-1033 [FREE Full text] [doi: 10.1016/j.chb.2010.03.003]
68. Dewey J. Democracy And Education: An Introduction To The Philosophy Of Education. New York City: Nabu Press; 1916.
69. YouTube. 2015. Fundacion Argentina de Transplante Hepatico. The man and the dog URL: https://www.youtube.com/watch?v=nySUhtaf2QQ [accessed 2019-10-08]
70. Health RA. Organ donation and transplantation: how does it work?. Organ donation and transplantation: how does it work?; 2013 Jan 01. URL: https://www.youtube.com/watch?v=HuKx2a5HkIM [accessed 2019-10-08]
71. Hargie O. The Handbook Of Communication Skills. Fourth Edition. New York: Routledge; 2018.
72. Spitzberg B, Cupach W. Interpersonal Communication Competence. Beverly Hills, CA: Sage Publications Inc; 1984.
73. Gordon E, Agu C, Siminoff L, Alolod G, Kim R, Gonzalez T, et al. Organ donation attitudes and behaviors among older hispanic women and hispanic community health workers. Transplantation 2017;101:S60. [doi: 10.1097/01.tp.0000525070.59547.4a]
74. Guidry JJ, Walker VD. Assessing cultural sensitivity in printed cancer materials. Cancer Pract 1999;7(6):291-296. [doi: 10.1046/j.1523-5394.1999.76005.x] [Medline: 10732526]
75. Kim S, Stoel L. Apparel retailers: website quality dimensions and satisfaction. J Retail Consum Serv 2004;11(2):109-117 [FREE Full text] [doi: 10.1016/S0969-6989(03)00010-9]

Abbreviations

HRSA/DoT: Health Resources and Services Administration's Division of Transplantation
ODM: Organ Donation Model

OMH: Office of Minority Health
TRA: Theory of Reasoned Action
VIT: Vested Interest Theory

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by A Waterman, N Yamashiki; comments to author 14.09.19; revised version received 30.09.19; accepted 01.10.19; published 13.01.20.

Please cite as:
Alolod GP, Gardiner H, Agu C, Turner JL, Kelly PJ, Siminoff LA, Gordon EJ, Norden R, Daly TA, Benitez A, Hernandez I, Guinansaca N, Winther LR, Bergeron CD, Montalvo A, Gonzalez T
A Culturally Targeted eLearning Module on Organ Donation (Promotoras de Donación): Design and Development
J Med Internet Res 2020;22(1):e15793
URL: https://www.jmir.org/2020/1/e15793
doi:10.2196/15793
PMID:31929102

Original Paper

Tools for App- and Web-Based Self-Testing of Cognitive Impairment: Systematic Search and Evaluation

Anna Pavlina Charalambous[1], PhD; Annie Pye[2], PhD; Wai Kent Yeung[2], BSc, MSc, PhD; Iracema Leroi[2,3], MD; Malcolm Neil[4]; Chryssoula Thodi[1], PhD; Piers Dawes[5], PhD

[1]School of Sciences, European University Cyprus, Nicosia, Cyprus

[2]Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester, United Kingdom

[3]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

[4]University of Dundee, Dundee, United Kingdom

[5]Manchester Centre for Audiology and Deafness, University of Manchester and the Manchester Academic Health Sciences Centre, Manchester, United Kingdom

Corresponding Author:
Piers Dawes, PhD
Manchester Centre for Audiology and Deafness
University of Manchester and the Manchester Academic Health Sciences Centre
A3.09 Ellen Wilkinson Building
Manchester, M13 9PL
United Kingdom
Phone: 44 1613061758
Email: piers.dawes@manchester.ac.uk

Abstract

Background: Tools for app- and Web-based self-testing for identification of cognitive impairment are widely available but are of uncertain quality.

Objective: The objective of this study was to undertake a scoping review of app- and Web-based self-tests for cognitive impairment and determine the validity of these tests.

Methods: We conducted systematic searches in electronic databases, including Google search, Google Play Store, and iPhone Operating System App Store, using the search terms “Online OR Internet-based AND Memory OR Brain OR Dementia OR mild cognitive impairment OR MCI AND Test OR Screen OR Check.”

Results: We identified 3057 tools, of which 25 were included in the review. Most tools meeting the inclusion criteria assessed multiple cognitive domains. The most frequently assessed domains were memory, attention, and executive function. We then conducted an electronic survey with the developers of the tools to identify data relating to development and validation of each tool. If no response to the survey was received, Google (to identify gray literature), Google Scholar, and Medical Literature Analysis and Retrieval System Online were searched using key terms “(name of developer, if available)” AND “(the name of the tool)” to identify any additional data. Only 7 tools had any information concerning psychometric quality, and only 1 tool reported data on performance norms, reliability, validity, sensitivity, and specificity for the detection of cognitive impairment.

Conclusions: The number of cognitive self-assessment electronic health tools for cognitive impairment is increasing, but most are of uncertain quality. There is a need for well-validated tools and guidance for users concerning which tools provide reliable information about possible cognitive impairment that could warrant further investigation.

(J Med Internet Res 2020;22(1):e14551) doi:10.2196/14551

KEYWORDS

telemedicine; eHealth; mHealth; dementia; mild cognitive impairment; self-assessment

Introduction

Background

By 2050, the number of people living with dementia is expected to increase to 152 million globally [1]. In many areas of the world, dementia remains underdiagnosed. For example, in many high-income countries, less than half of the people living with dementia receive a formal diagnosis [2]. In low- and middle-income countries (LMICs), less than 10% of the people with dementia receive a formal diagnosis [1,3-5]. Raising awareness of dementia and self-detection of cognitive decline have the potential to increase the diagnosis rate, thereby fostering appropriate support for people with dementia. Self-detection may also facilitate earlier diagnosis, a critical aspect of dementia care [6]. Several national policies relating to dementia have prioritized the need to increase the diagnosis rate and ensure a timely diagnosis of dementia [6]. Self-testing for cognitive impairment, via app- or Web-based tools, may support these aspirations by identifying people who may be developing cognitive impairment and by directing them to appropriate health and social care support services.

Electronic medical and mental health information and services (referred to as electronic health [*eHealth*]) are increasingly delivered through the internet [7]. eHealth tools include app- and Web-based tools that can screen individuals who are at risk and/or offer self-help intervention or clinical referrals for various health conditions [8]. Dramatic increases in access to and use of internet and mobile phone technology have supported the development of eHealth technology. In 1995, only 1% of the global population had an internet connection [9]. In 2019, this percentage had increased to 56.3% [10]. The majority of the worldwide population that previously did not have access to a computer or a fixed-line telephone now has mobile phones [11]. In Europe, there has been a 583% growth in internet usage between the years 2000 and 2019; over 85% of the population now has access to the internet [10].

The potential benefits of eHealth apps include improvements in health safety, improved health care efficiency and effectiveness, reduced costs, improved decision making (eg, in reaching a diagnosis), access to remote clinicians, and medical error reduction [12]. The potential benefits of eHealth apps are substantial, but there is also potential for harm. The quality, safety, and effectiveness of the majority of the proliferating eHealth apps are unknown (eg, study by Eng and Lee [13]). Health professionals have often not been involved in the development of eHealth tools, and the tools have frequently been developed without appropriate validation [14-17]. Uncertainty in the quality of eHealth tools is worrisome in relation to tools for self-identification of cognitive impairment indicative of dementia. Furthermore, if not properly validated, there is a risk of false-positive identification that may cause needless anxiety or false-negative identification that can result in a diagnosis of dementia being missed. Formal studies are urgently needed to establish the potential benefits and mitigate the harms of mobile health (mHealth) technology.

Objectives

This review aimed to identify and assess (1) the numbers, availability, and characteristics of app- and Web-based self-assessment tools for cognitive impairment and (2) the psychometric quality of these tools to inform their future development for self-assessment of cognitive impairment.

Methods

Design

A systematic search was conducted between May 2017 and May 2018 by researchers at the University of Manchester, United Kingdom. We identified Web-based tools through the Google search engine and mobile phone and tablet apps through Google Play and the iPhone Operating System (iOS) App Store. The search terms we used were “Online OR Internet-based AND Memory OR Brain OR Dementia OR mild cognitive impairment OR MCI AND Test OR Screen OR Check.” We searched the iOS App Store and Google Play using the same search terms as in Google search, with the exception of the “online” and “internet-based” search terms. We screened the first 100 results we identified in each search for relevance according to the inclusion or exclusion criteria. Around 75% of *clicks* are in relation to the first 20 hits obtained [18]. We evaluated the first 100 results as a liberal criterion to capture all the research results that users would likely encounter. We completed a follow-up search in November 2018 and a further follow-up search in February 2019 to check that the tools identified in the first search were still available on the Web.

Inclusion and Exclusion Criteria

A tool was suitable for inclusion if it (1) was designed to be a self-administered cognitive tool, (2) was hosted on the Web or as a mobile phone app, (3) was a tool intended for detection and/or assessment of (all cause) dementia and/or mild cognitive impairment (MCI), and (4) is available for free or at a low cost (≤£5). A sum of £5 was chosen as an upper limit for cost to select tools that are readily available and are within the average price range for mobile apps [19]. Games; puzzles; *brain training* apps; IQ, vocational, or academic achievement tools; tools for specific learning disability (eg, dyslexia); or tools that estimate future risk of dementia based on lifestyle factors were excluded as these tools do not measure current cognitive ability.

The first reviewer screened all titles obtained through each search, identifying candidate tests for inclusion. The second reviewer then screened 10% of the titles to ensure consensus opinion. Both reviewers held PhDs in cognitive psychology. Both reviewers evaluated all tools identified against the inclusion criteria, and in instances of disagreement, the third reviewer acted as an arbitrator and decided whether the tool met the criteria for inclusion.

Survey of Psychometric Quality

We identified a point of contact for the owner or developer of each tool from the respective apps or websites and sent an email invitation to complete an electronic survey about the tool. The survey was adapted from a postal survey of tests or batteries for assessment of MCI [20]. We obtained permission to adapt

the survey from the author of the original version. The survey contained questions about the content of the tool (ie, the cognitive domains it assesses), the duration of the test, normative data, and whether validity and reliability have been established (see Multimedia Appendix 1).

We collected the survey data over a 3-month period. In week 1, we sent out a covering email and survey link to the points of contact identified for each tool. In the subsequent 4 weeks, we sent weekly follow-up emails to those who had not responded and fortnightly reminder emails thereafter. Those who had not responded after 12 weeks were not contacted any further. For those tools for which we received no response to the survey or for which we could not identify and/or contact the owner of the tool, we conducted supplementary searches. Supplementary searches were run on Google (for gray literature), Google Scholar, and Medical Literature Analysis and Retrieval System Online (for published or peer-reviewed articles) to identify information relating to the development and validation of each tool. The search terms used were the name of developer (if available) AND the name of the tool. We screened the first 100 results, sorted by relevance. We then downloaded relevant titles and Web pages and saved the Web page links or relevant papers. Finally, we extracted data relating to the development and validation of each tool from the material we had identified in supplementary searches using a data entry table based on the same parameters of the survey questionnaire.

Results

Search Findings

We identified 3057 tools (apps and websites) after searching Google search (n=1205), Google Play (n=1201), and iOS App Store (n=651; Figure 1). The initial search identified 39 tests that fit the inclusion criteria. After a follow-up search in May 2018, we removed 3 tools after we found that they had been removed from their respective app stores, leaving us with 36 tools. We sent the survey to 32 out of the 36 tools identified. For 4 tools, we were unable to identify a contact email or emails could not be delivered, and so we were unable to contact the test developers. After the survey had been completed, 1 extra tool was added as a test developer indicated its existence, resulting in 37 tools. Follow-up searches revealed that 12 tools had been removed from their app stores, leaving 25 tools included in the review (Table 1). After sending out the online survey, we obtained responses for 9 tools. Subsequently, we obtained information for 4 additional tools through supplementary searches (see Figure 2).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram indicating the app or tools search and screening process.

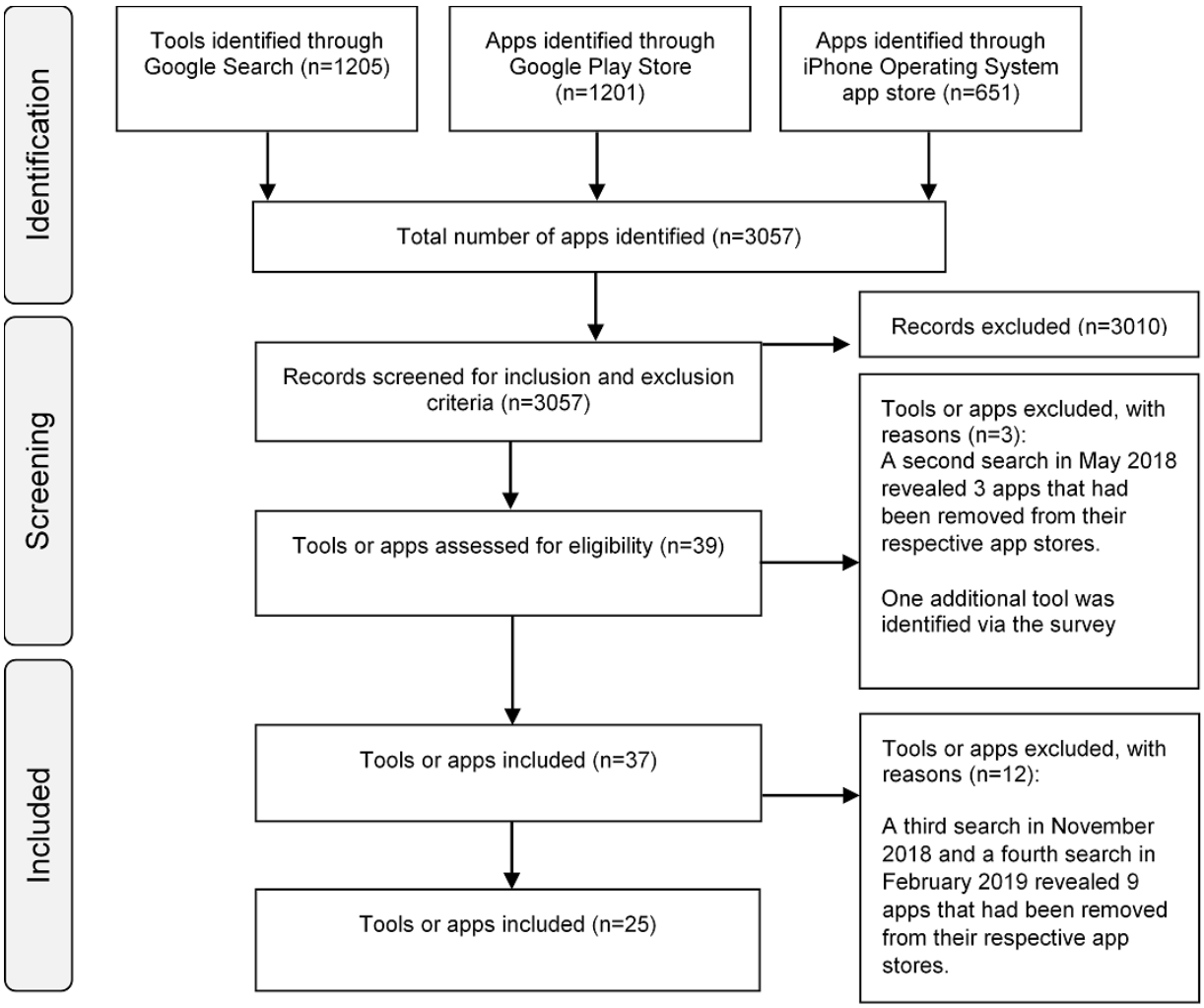


Table 1. Cognitive domains assessed by each app- and Web-based cognitive tool.

Test name	Platform	Attention	Memory	Executive function	Visual spatial ability	Other	Not available
BrainTest (electronic Self-Administered Gerocognitive Examination)	1[a] and 2[b]	✓[c]	✓	✓	✓	✓	—[d]
BrainCheck	1	✓	✓	✓	✓	—	—
MemTrax–The Online Memory Screening Test (free version)	1	✓	✓	—	—	✓	—
MemTrax Proprietary	1	✓	✓	✓	✓	✓	—
Self-Assessment of Cognition	1	—	✓	✓	—	—	—
Husketest	1 and 3[e]	—	✓	—	—		—
Dementia Screener	2	✓	✓	✓	—	✓	—
DANA[f] Brain Vital	2 and 3	—	—	—	—	✓	—
DANA Modular	2 and 3	—	—	—	—	✓	—
Cogniciti	1	✓	✓	—	—	—	—
Savonix Mobile	2	✓	✓	✓	—	✓	—
Imprint Memory Assessment	1	—	✓	—	—	—	—
Memory Quiz	1	—	✓	—	—	—	—
Dementia Test	1	—	✓	—	—	—	—
RateMyMemory	1	—	✓	—	—	—	—
Daily Mail Dementia Quiz	1	—	✓	—	—	—	—
Cognitive Function Test	1	—	—	—	—	—	✓
The Cleveland Clinic Brain Check-Up	1	—	✓	—	—	—	—
Mindcrowd	1	—	✓	—	—	—	—
MyBrainTest	1	—	✓	—	—	—	—
Memory Health Check	1	—	✓	—	—	—	—
On Memory	1	—	✓	—	—	—	—
Psychology Today Memory Test	1	—	✓	—	—	—	—
Brainlab Cognition	3	—	—	—	—	—	✓
Dementia Test–Risk Calculator of Dementia	3	✓	✓	—	—	✓	—
MMSE[g]	2	—	—	—	—	—	✓
Total	—	8	21	6	2	7	3

[a]1 signifies Google search.

[b]2 signifies Google Play Store.

[c]Assessed domain.

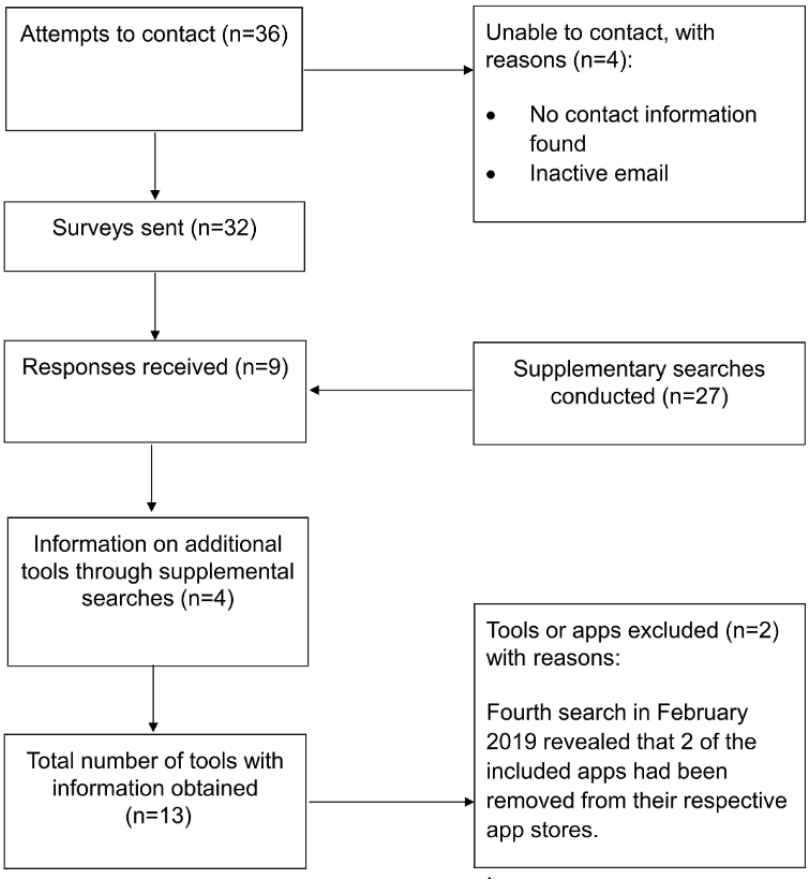
[d]Data not assessed/data not available.

[e]3 signifies iPhone Operating System App Store.

[f]DANA: Defense Automated Neurobehavioral Assessment.

[g]MMSE is not related to the Mini–Mental State Exam [20]

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram describing the Web-based survey and supplemental searches.



We identified 17 of the 25 tools from searching Google search, 2 by searching the iOS platform, 4 by searching Google Play, 1 was identified both in Google search and Google Play, and 1 in Google Play and iOS platform (Table 1). The time of test completion was reported to range from 1.5 min to 30 min. One of the Web-based cognitive tools (BrainTest [Electronic Self-Administered Gerocognitive Examination; eSAGE] tool [21]) was the digital version of the Self-Administered Gerocognitive Examination (SAGE) [22]. Most of the tools purported to assess multiple cognitive domains (Table 1). The most frequently assessed cognitive domain was memory (tested by 21 tools), then attention (8 tools), followed by executive function (6 tools). Cognitive domains that were less frequently assessed (the *Other* category) included orientation, language, fluency, and reaction time. If the cognitive domains tested by an included tool were not explicitly stated or could not be identified by reading the instructions of the tool, it was reported in Table 1 as not available.

A total of 6 survey respondents provided information concerning the collection of normative data, reliability, validity, and sensitivity/specificity for the detection of cognitive impairment (Table 2). We identified psychometric data only for 1 additional test in supplementary searches. For the rest of the tests, no psychometric data were available.

Table 2. Summary of availability of psychometric test data.

Test name	Normative data	Reliability	Validity	Sensitivity and specificity[a]
BrainTest (electronic Self-Administered Gerocognitive Examination) [21]	✓[b]	✓✓[c]	✓	✓✓
BrainCheck [23]	✓✓	✓✓	✓✓	×[d]
MemTrax–The Online Memory Screening Test (free version) [24]	×	×	×	×
MemTrax Proprietary [24]	✓✓	✓	✓	×
Self-Assessment of Cognition [25]	✓	✓	×	×
Husketest [26,27]	✓✓	×	✓✓	×
Dementia Screener [28]	×	×	×	×
DANA[e] Brain Vital [29]	✓✓	✓✓	✓✓	✓✓
DANA Modular [30]	✓✓	✓✓	✓✓	✓✓
Cogniciti [31]	✓✓	✓✓	✓✓	×
Savonix Mobile [32]	×	×	×	×
Imprint Memory Assessment [33]	×	×	×	×
Memory Quiz [34]	×	×	×	×
Dementia Test [35]	×	×	×	×
RateMyMemory [36]	×	×	×	×
Daily Mail Dementia Quiz [37]	×	×	×	×
Cognitive Function Test [38]	×	×	×	×
The Cleveland Clinic Brain Check-Up [39]	×	×	×	×
Mindcrowd [40]	×	×	×	×
MyBrainTest [41]	×	×	×	×
Memory Health Check [42]	×	×	×	×
On Memory [43]	×	×	×	×
Psychology Today Memory Test [44]	×	×	×	×
Brainlab Cognition [45]	×	×	×	×
Dementia Test–Risk Calculator of Dementia [46]	×	×	×	×
MMSE[f] [47]	×	×	×	×

[a]To detect dementia or mild cognitive impairment.

[b]One tick indicates data reported to be in preparation.

[c]Two ticks indicate data currently available.

[d]A cross indicates no data available or no response.

[e]DANA: Defense Automated Neurobehavioral Assessment.

[f]MMSE is not related to the Mini–Mental State Exam [20]

BrainTest (Electronic Self-Administered Gerocognitive Examination)

BrainTest is based on the SAGE [22], a brief cognitive assessment for identification of MCI and early dementia. Developers reported that they expected to have normative data by the end of 2019. Spearman correlations between eSAGE with SAGE (r_s=0.88), Montreal Cognitive Assessment (MoCA; r_s=0.76), and Mini-Mental State Examination (MMSE; r_s=0.67) were strong, demonstrating high convergent validity. The developers found no difference between the eSAGE and SAGE with regard to sensitivity or specificity in differentiating people without dementia (MCI or normal) from those with dementia [48]. The sensitivity and specificity of eSAGE were 90% and 87%, respectively, for differentiating dementia from normal cognition and were 90% and 75%, respectively, for differentiating MCI from dementia [48]. eSAGE reportedly had 90% specificity and 71% sensitivity in differentiating those with cognitive impairment (MCI and dementia) from those with normal cognitive function [48]. The developers of BrainTest reported that data on the test-retest reliability of eSAGE were available; however, they did not share or identify any information on reliability.

BrainCheck

BrainCheck offers a collection of neurocognitive tests intended to track cognitive health over time. According to its developers, BrainCheck has a normative database that contains more than 20,000 test results, but they did not provide any further details. BrainCheck had high 7-day test-retest reliability, with correlation coefficients ranging from 0.6 to 0.9 by subtest. The BrainCheck website reported data on sensitivity and specificity of identifying traumatic brain injury versus normal cognition [49], but no data with regard to dementia or MCI were reported. BrainCheck developers reported that a publication was underway reporting the validity of BrainCheck in relation to dementia.

MemTrax

MemTrax is a test of recognition memory that is intended for early detection of memory problems that may be indicative of dementia. Normative data are reportedly available for the proprietary version of the MemTrax test. In a 2011 validation study, the developers gathered data from 868 individuals from 25 sites (including community events, senior citizen centers, and retirement living communities in the San Francisco Bay Area). The age range of participants was 40 to 97 years; 68.7% were female with formal education ranging from 6 to 21 years [50]. Recognition memory declined with age, and the decline was accompanied with a greater than 3-fold increase in variability over the age range. Individuals with more than 13 years of education had higher scores than those with fewer years of formal education [50]. The developers reported that test-retest reliability and convergent validity data for MemTrax (vs MoCA) [51] were being collected.

Self-Assessment of Cognition

Self-Assessment of Cognition (SAC) is a brief cognitive screening tool designed to give older adults and their health care professionals information about memory and cognitive functioning. The developers reported that SAC has normative data from a combined sample of 206 residents of long-term care facilities for older adults (manuscripts for both studies were reportedly in preparation). The developers reported that they are collecting reliability and validity data.

Husketest

Husketest is a multiple-choice picture recognition test. The developers of Husketest reported that they collected normative data from 795 individuals with an age range of 4 to 86 years. They also reported small effects of education and age on performance and that the test suffers from a ceiling effect, which may limit the sensitivity of the test. The developers provided no further details.

Defense Automated Neurobehavioral Assessment Modular

This US Food and Drug Administration (FDA)–approved tool comprises a suite of 8 cognitive tests and 7 psychological surveys. The developers gathered normative data from 814 adult military veterans (71% male) aged between 18 and 64 years [52]. The test-retest reliability of Defense Automated Neurobehavioral Assessment (DANA) subtests procedural reaction time (PRT) and simple reaction time was 0.75 and 0.81, respectively [52]. The developers assessed the sensitivity of DANA in detecting MCI and Alzheimer Disease (AD) in a pilot study with 7 patient and caregiver dyads [53]. The group with AD or cognitive impairment performed worse than the caregivers for all the subtests of DANA, apart from simple reaction time [53]. Finally, the developers reported relationships among 3 DANA subtests, namely, memory search, PRT, and spatial processing with MMSE scores [54], but they did not report the correlation values.

Cogniciti

Cogniciti is a self-assessment tool that is intended to be used by individuals to determine whether they should raise their concerns about memory with their primary care provider. Cogniciti includes (1) spatial working memory, (2) Stroop interference, (3) face-name association, and (4) letter-number alteration subtests. The developers collected normative data from 361 healthy adults aged 50 to 79 years [55]. Internal consistency (Cronbach alpha) of the face-name association test was below acceptable levels (alpha=.62), whereas consistency was excellent (alpha=.96) for the Stroop interference subtest. There were insufficient trials to calculate the internal consistency for the other 2 tasks. The developers reported retest-reliability ranging between 0.49 and 0.82 for individual subtests and 0.72 for the overall score. The developers calculated correlations between age and performance for each subtest as a measure of construct validity. These correlations were small to medium in size: −0.20 to 0.31. They examined intersubtest correlations as a measure of convergent validity, and these correlations were again small to medium in size: −0.27 to 0.30.

Discussion

Principal Findings

This is the first review of the quantity and quality of app- and Web-based self-assessment tools for cognitive impairment. We identified 25 tools via Google search, Google Play, and iOS platform searches, but only 7 tests had any information concerning psychometric quality, and only 1 tool (DANA) reported data on performance norms, reliability, validity, and sensitivity or specificity for the detection of cognitive impairment. The lack of information about the psychometric properties of the majority of tools indicates that although the number of cognitive self-assessment eHealth tools is increasing [11,56], their quality is unknown. The uncertain validity of the majority of tools is a concern as some tools may fail to identify people who have cognitive impairment or may cause undue anxiety by falsely identifying cognitive impairment. The focus of this review was on self-assessment tools that were not intended to inform or replace clinical decisions nor could be used to inform provision of treatment, that is, are not medical devices. However, the definition of software as a medical device is unclear, particularly in the context of software intended to identify cognitive impairment indicative of dementia. Classification of software as a medical device should consider the potential of the software to cause harm (in this case, by falsely identifying or missing true cases of cognitive impairment). We suggest that it is important to establish standards and identify ways of conveying the reliability of tools

to users so that users are able to make informed choices about the tools they use and the results obtained from each tool. Developers have an ethical duty to establish the psychometric quality of the tools they offer and provide appropriate caveats on the interpretation of results obtained as well as give specific instructions for acting on the results of the self-test. eHealth apps tend to be categorized in the *Health and Fitness* or *Medical* sections of app stores. This terminology may encourage users to view these tools as legitimate sources of medical information. Formal regulation by national authorities (eg, the US FDA or the UK Medicines and Healthcare products Regulatory Agency) is a possibility, although they may struggle to keep pace with the rapid development in eHealth [57]. The FDA recently proposed that it would regulate only those apps that provide diagnostic and treatment recommendations to physicians [58,59]. This new guidance excludes all app- and Web-based self-assessment tools [60].

In the European Union's model, developers can file an application for medical device registration with any member state of the European Union. The Conformité Européenne mark issued by the respective body in each member state is then valid throughout the European Union. The European regulatory system could offer another potential model for regulation. However, medical devices approved in Europe only need to establish performance and safety but not clinical efficacy or effectiveness [60]. The 1988 Clinical Laboratory Improvement Amendments (CLIA) model has been suggested as a possible solution for the regulation of mHealth, including mobile computing, medical sensor, and communication technologies for health care apps [8,61]. CLIA is a system for ensuring that diagnostic testing laboratories comply with US regulatory standards. Nonprofit accrediting agencies with authority to issue certification under federal CLIA standards ensure consistency of record keeping and staff training [61,62]. Larson [60] argued that a CLIA model could ensure that mHealth apps comply with basic standards, including (1) accessibility: clear language, usability, and affordability; (2) privacy and security, including data sharing with third parties and; (3) content: apps developed with health care experts contain accurate information, limit advertising, and explain monetization (eg, referral generation or sales) and conflicts of interest [17]. Existence of a recognized standard of quality for mHealth apps could provide an additional incentive for developers to establish the psychometric quality of the tools they provide. Establishing the psychometric quality of assessments requires significant investment. Developing and establishing the psychometric properties of a cognitive test requires psychometric expertise and carefully controlled testing of large numbers of people with and without cognitive impairment. If an app is made available for free or at a low cost, developers would have to either (1) have a business model that funds development and running costs without directly transferring those costs to the end user (eg, apps that make referrals to for-profit health care providers) or (2) be well funded by charity, social enterprise, or government organizations (eg, DANA developed by the US military). Development cost limitations are likely to limit the number of good quality self-assessment apps.

The tools in this review were all either free or low cost and readily accessible. eHealth tools may have the potential to address the underidentification of dementia by increasing the awareness of cognitive impairment and directing people who may have cognitive difficulties to appropriate clinical diagnostic and support services. Overall, 2 tools mentioned in this review were linked to clinical services (Cogniciti) or provided specific recommendations to speak with a doctor (eSAGE). Furthermore, 3 tests provided advice on healthy lifestyles to promote cognitive health (Daily Mail Dementia Quiz, Memory Health Check, MyBrainTest, and On Memory). However, 19 out of the 25 tools included in the review did not directly link users with a clinician or support service. The lack of clear advice or direction on what steps should be taken in the event of a failed screening is of concern; some tests may cause anxiety by identifying a possible impairment without providing advice about how to seek help. Lack of clear direction may also mean that few people may act on the results of a failed cognitive screening. We did not identify any study that evaluated the proportion of people who failed the screening or who went on to seek help. In addition, none of the studies identified barriers to acting on the results of failed tests nor investigated how help might best be provided. The lack of follow-up actions may be a serious shortcoming of most Web-based cognitive self-assessments.

Further efforts should be made to link the results of Web-based cognitive self-assessments directly to clinical services to minimize anxiety caused by identification of a possible impairment and facilitate action on the result of a failed screening. eHealth cognitive self-assessment tools could potentially utilize the growing acceptance and use of video conferencing in geriatric psychiatry care by clinicians and patients. Videoconferencing is well received by patients and clinicians [63] and may facilitate the reach of clinical services in underserved areas [64]. Video consultations may help increase diagnostic coverage particularly in LMICs. As an alternative to referral to clinical services, Web-based cognitive assessments could direct test takers to interventions delivered via the internet to support cognitive function. For example, the Imprint Memory Assessment eHealth tool links to a Web-based memory health program based on diet, exercise, cognitive training, and vascular risk monitoring [65]. Formal evaluation of the risks and benefits of a cognitive eHealth paradigm is an urgent priority.

Conclusions

There is a need to establish the quality of cognitive self-assessment tools while maintaining their low cost and easy accessibility. A regulatory model should ensure standards of accessibility, privacy, and content. The results of app- and Web-based cognitive self-assessments should be linked to appropriate clinical and support services.

Acknowledgments

This work was supported by SENSE-Cog project. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 668648. PD is supported by the National Institute for Health Research Manchester Biomedical Research Centre.

Authors' Contributions

APC conducted an update of the searches and wrote the manuscript, including the Introduction, Methods, Results, and Discussion. PD contributed to the development of the search strategy and methodology and provided significant input to the manuscript, including the Introduction, Methods, Results, and Discussion sections. AP contributed to the development of the search strategy and methodology and conducted the initial searches with MN. WKY contributed to the development of the manuscript. IL is the chief investigator for the SENSE-Cog project and contributed to the development of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOC File , 77 KB - jmir_v22i1e14551_app1.doc]

References

1. Patterson C. Alzheimer's Disease International. London: Alzheimer's Disease International; 2018 Sep. World Alzheimer Report 2018 - The State of the Art of Dementia Research: New Frontiers URL: https://www.alz.co.uk/research/WorldAlzheimerReport2018.pdf [accessed 2019-12-03]
2. Waldemar G, Phung KT, Burns A, Georges J, Hansen FR, Iliffe S, et al. Access to diagnostic evaluation and treatment for dementia in Europe. Int J Geriatr Psychiatry 2007 Jan;22(1):47-54. [doi: 10.1002/gps.1652] [Medline: 17044135]
3. Nakamura AE, Opaleye D, Tani G, Ferri CP. Dementia underdiagnosis in Brazil. Lancet 2015 Jan 31;385(9966):418-419. [doi: 10.1016/S0140-6736(15)60153-2] [Medline: 25706975]
4. Jitapunkul S, Chansirikanjana S, Thamarpirat J. Undiagnosed dementia and value of serial cognitive impairment screening in developing countries: a population-based study. Geriatr Gerontol Int 2009 Mar;9(1):47-53. [doi: 10.1111/j.1447-0594.2008.00501.x] [Medline: 19260979]
5. Dias A, Patel V. Closing the treatment gap for dementia in India. Indian J Psychiatry 2009 Jan;51(Suppl 1):S93-S97 [FREE Full text] [Medline: 21416026]
6. Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagiannidou M. Alzheimer's Disease International. London: Alzheimer's Disease International; 2016 Sep. World Alzheimer Report 2016. Improving Healthcare for People Living With Dementia: Coverage, Quality, and Costs Now and in the Future URL: https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf [accessed 2019-12-03]
7. World Health Organization, International Telecommunication Union. National eHealth Strategy Toolkit. Geneva: World Health Organization; 2012.
8. Mechael PN. The case for mHealth in developing countries. Innov Technol Governance Glob 2009;4(1):103-118. [doi: 10.1162/itgg.2009.4.1.103]
9. Internet Live Stats. 2016. Number of Internet Users URL: http://www.internetlivestats.com/internet-users/ [accessed 2019-02-12] [WebCite Cache ID 781TVhU6C]
10. Internet World Stats - Usage and Population Statistics. 2019. World Internet Users and 2019 Population Stats URL: http://www.internetworldstats.com/stats.htm [accessed 2019-04-16] [WebCite Cache ID 781TmbfPd]
11. Ivatury G, Moore J, Bloch A. A doctor in your pocket: health hotlines in developing countries. Innov Technol Governance Glob 2009 Jan;4(1):119-153. [doi: 10.1162/itgg.2009.4.1.119]
12. Botha M, Botha A, Herselman M. The Benefits and Challenges of e-Health Applications: A Content Analysis of the South African context. In: Proceedings of The International Conference on Computer Science, Computer Engineering, and Social Media. 2014 Presented at: CSCESM'14; December 12-14, 2014; Thessaloniki, Greece.
13. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
14. Hamilton AD, Brady RR. Medical professional involvement in smartphone 'apps' in dermatology. Br J Dermatol 2012 Jul;167(1):220-221. [doi: 10.1111/j.1365-2133.2012.10844.x] [Medline: 22283748]
15. 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]
16. Rodrigues MA, Visvanathan A, Murchison JT, Brady RR. Radiology smartphone applications; current provision and cautions. Insights Imaging 2013 Oct;4(5):555-562 [FREE Full text] [doi: 10.1007/s13244-013-0274-4] [Medline: 23912880]

17. Cook VE, Ellis AK, Hildebrand KJ. Mobile health applications in clinical practice: pearls, pitfalls, and key considerations. Ann Allergy Asthma Immunol 2016 Aug;117(2):143-149. [doi: 10.1016/j.anai.2016.01.012] [Medline: 27499542]
18. Shelton K. Forbes Magazine. 2017 Oct 30. The Values of Search Results Rankings URL: https://www.forbes.com/sites/forbesagencycouncil/2017/10/30/the-value-of-search-results-rankings/#66ffa9f644d3 [accessed 2019-12-03]
19. Statista. Average prices for apps in the Apple App Store as of November 2019 URL: https://www.statista.com/statistics/267346/average-apple-app-store-price-app/ [accessed 2019-02-22] [WebCite Cache ID 781TxyIZp]
20. Snyder PJ, Jackson CE, Petersen RC, Khachaturian AS, Kaye J, Albert MS, et al. Assessment of cognition in mild cognitive impairment: a comparative study. Alzheimers Dement 2011 May;7(3):338-355 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.009] [Medline: 21575877]
21. BrainTest (eSAGE). URL: https://braintest.com [accessed 2019-04-30] [WebCite Cache ID 781BvQDqE]
22. Scharre DW, Chang S, Murden RA, Lamb J, Beversdorf DQ, Kataki M, et al. Self-administered Gerocognitive Examination (SAGE): a brief cognitive assessment Instrument for mild cognitive impairment (MCI) and early dementia. Alzheimer Dis Assoc Disord 2010;24(1):64-71. [doi: 10.1097/WAD.0b013e3181b03277] [Medline: 20220323]
23. WebCite. BrainCheck URL: https://www.webcitation.org/781JpWIPq [accessed 2019-04-30] [WebCite Cache ID 781JpWIPq]
24. MemTrax. URL: https://memtrax.com/ [accessed 2019-04-30] [WebCite Cache ID 781M2Z93g]
25. WebCite. Self-Assessment of Cognition URL: https://www.webcitation.org/781MuD8zc [accessed 2019-04-30] [WebCite Cache ID 781MuD8zc]
26. iTunes - Apple. Husketest App URL: https://itunes.apple.com/is/app/husketest/id853661406?mt=8 [accessed 2019-04-30]
27. National Knowledge Center for Dementia. Memory test - test your memory URL: http://www.videnscenterfordemens.dk/husketest [accessed 2019-04-30] [WebCite Cache ID 781MwODDy]
28. Google Play Store. Dementia Screener URL: https://play.google.com/store/apps/details?id=appinventor.ai_lcchien86.DementiaScreener [accessed 2019-04-30]
29. Google Play Store. DANA Brain Vital URL: https://play.google.com/store/apps/details?id=com.atinc.staging.brainVital&hl=en_GB [accessed 2019-04-30]
30. Google Play Store. DANA Modular URL: https://play.google.com/store/apps/details?id=com.atinc.staging.full&hl=en_GB [accessed 2019-04-30]
31. Cogniciti. URL: https://cogniciti.com/ [accessed 2019-04-30] [WebCite Cache ID 781MzGauE]
32. Google Play Store. Savonix Mobile URL: https://play.google.com/store/apps/details?id=com.savonix.savonixbeta&hl=en_GB&showAllReviews=true [accessed 2019-04-30]
33. Neurotrack. URL: https://neurotrack.com/ [accessed 2019-04-30] [WebCite Cache ID 781N1Q4q5]
34. Alzheimer's Prevention. Memory Quiz URL: http://alzheimersprevention.org/alzheimers-info/memory-quiz/ [accessed 2019-04-30] [WebCite Cache ID 781N3DKxD]
35. IllnessQuiz. Dementia Test URL: https://illnessquiz.com/dementia-test/ [accessed 2019-04-30] [WebCite Cache ID 781N56bXU]
36. Memory Rate. URL: http://memoryrate.com/ [accessed 2019-04-30] [WebCite Cache ID 781N6NVky]
37. Hodgekiss A. Daily Mail Online. 2017 Mar 20. Do You Have a Memory Like a Goldfish? Take This Test to Find Out if You or Your Loved Ones Have a Reason to Worry URL: https://www.dailymail.co.uk/health/article-4332608/Do-memory-like-goldfish.html [accessed 2019-04-30] [WebCite Cache ID 781N8Bf1N]
38. Food for the Brain. Cognitive Function Test URL: https://www.foodforthebrain.org/alzheimers-prevention/take-the-test.aspx [accessed 2019-04-30] [WebCite Cache ID 781NA61Fp]
39. Healthy Brains by Cleveland Clinic. Get Your Free Online Self Assessment URL: https://healthybrains.org/brain-check-up/ [accessed 2019-04-30] [WebCite Cache ID 781NBhFrz]
40. MindCrowd. URL: https://www.mindcrowd.org/ [accessed 2019-04-30] [WebCite Cache ID 781NDwY4R]
41. MyBrainTest. Online Memory Tests URL: https://www.mybraintest.org/online-memory-screening-tests/ [accessed 2019-04-30] [WebCite Cache ID 781NFEiDJ]
42. MemoryHealthCheck. Free Memory Loss Tests URL: https://www.memorylosstest.com/free-memory-loss-tests-online/ [accessed 2019-04-30] [WebCite Cache ID 781NGskHv]
43. On Memory. Signs & Symptoms: Memory Test URL: https://www.onmemory.ca/en/signs-symptoms/memory-test [accessed 2019-04-30] [WebCite Cache ID 781NIAzIG]
44. Psychology Today. Memory Test URL: http://psychologytoday.tests.psychtests.com/bin/transfer?req=MTF8MzM2MHwxOTQyMjE5OHwwfDE=&refempt=1501665258.19422198.11 [accessed 2019-04-30] [WebCite Cache ID 781NM3zfm]
45. App Store - Apple. Brainlab Cognition URL: https://itunes.apple.com/gb/app/brainlab-cognition/id1109843085?mt=8 [accessed 2019-04-30]
46. App Store - Apple. Dementia Test - Risk Calculator of Dementia URL: https://itunes.apple.com/gb/app/dementia-test-risk-calculator-of-dementia/id1014958634?mt=8 [accessed 2019-04-30]
47. Google Play Store. MMSE URL: https://play.google.com/store/apps/details?id=com.yasintanriverdi.mmse&hl=en_GB [accessed 2019-04-30]

48. Scharre DW, Chang SI, Nagaraja HN, Vrettos NE, Bornstein RA. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. Alzheimers Res Ther 2017 Jun 27;9(1):44 [FREE Full text] [doi: 10.1186/s13195-017-0269-3] [Medline: 28655351]
49. Yang S, Flores B, Magal R, Harris K, Gross J, Ewbank A, et al. Diagnostic accuracy of tablet-based software for the detection of concussion. PLoS One 2017;12(7):e0179352 [FREE Full text] [doi: 10.1371/journal.pone.0179352] [Medline: 28686616]
50. Ashford JW, Gere E, Bayley PJ. Measuring memory in large group settings using a continuous recognition test. J Alzheimers Dis 2011;27(4):885-895. [doi: 10.3233/JAD-2011-110950] [Medline: 21908910]
51. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
52. Russo CR, Lathan CE. An evaluation of the consistency and reliability of the defense automated neurocognitive assessment tool. Appl Psychol Meas 2015 Oct;39(7):566-572 [FREE Full text] [doi: 10.1177/0146621615577361] [Medline: 29881027]
53. Lathan CE, Coffman I, Shewbridge R, Lee M, Cirio R. A pilot to investigate the feasibility of mobile cognitive assessment of elderly patients and caregivers in the home. J Geriatr Palliat Care 2016;4(1):6. [doi: 10.13188/2373-1133.1000017]
54. Hollinger KR, Woods SR, Adams-Clark A, Choi SY, Franke CL, Susukida R, et al. Defense automated neurobehavioral assessment accurately measures cognition in patients undergoing electroconvulsive therapy for major depressive disorder. J ECT 2018 Mar;34(1):14-20. [doi: 10.1097/YCT.0000000000000448] [Medline: 28991066]
55. Troyer AK, Rowe G, Murphy KJ, Levine B, Leach L, Hasher L. Development and evaluation of a self-administered on-line test of memory and attention for middle-aged and older adults. Front Aging Neurosci 2014;6:335 [FREE Full text] [doi: 10.3389/fnagi.2014.00335] [Medline: 25540620]
56. Rietjens JA, Sudore RL, Connolly M, van Delden JJ, Drickamer MA, Droger M, European Association for Palliative Care. Definition and recommendations for advance care planning: an international consensus supported by the European Association for Palliative Care. Lancet Oncol 2017 Sep;18(9):e543-e551. [doi: 10.1016/S1470-2045(17)30582-X] [Medline: 28884703]
57. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
58. Food and Drug Administration. Maryland, USA: Food and Drug Administration; 2015. Policy for Device Software Functions and Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications [accessed 2019-04-30]
59. Food and Drug Administration. Maryland, USA: Food and Drug Administration; 2019 Sep. Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm587819.pdf [accessed 2019-04-30]
60. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
61. Centers for Medicare & Medicaid Services. 1988. Clinical Laboratory Improvement Amendments URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html [accessed 2019-03-14]
62. Yost JA. Laboratory inspection: the view from CMS. Lab Med 2003;34(2):136-140. [doi: 10.1309/uv7dpf0u581y2m8h]
63. Barton C, Morris R, Rothlind J, Yaffe K. Video-telemedicine in a memory disorders clinic: evaluation and management of rural elders with cognitive impairment. Telemed J E Health 2011 Dec;17(10):789-793. [doi: 10.1089/tmj.2011.0083] [Medline: 22023458]
64. Bossen AL, Kim H, Williams KN, Steinhoff AE, Strieker M. Emerging roles for telemedicine and smart technologies in dementia care. Smart Homecare Technol Telehealth 2015;3:49-57 [FREE Full text] [doi: 10.2147/SHTT.S59500] [Medline: 26636049]
65. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun 6;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]

Abbreviations

AD: Alzheimer Disease
CLIA: Clinical Laboratory Improvement Amendments
DANA: Defense Automated Neurobehavioral Assessment
eHealth: electronic health
eSAGE: Electronic Self-Administered Gerocognitive Exam
FDA: Food and Drug Administration
iOS: iPhone Operating System
LMIC: low- and middle-income country

MCI: mild cognitive impairment
mHealth: mobile health
MoCA: Montreal Cognitive Assessment
MMSE: Mini-Mental State Examination
PRT: procedural reaction time
SAC: Self-Assessment of Cognition
SAGE: Self-Administered Gerocognitive Examination

Edited by G Eysenbach; submitted 02.05.19; peer-reviewed by D Scharre, RK B, S Sankaran; comments to author 01.07.19; revised version received 16.08.19; accepted 24.09.19; published 17.01.20.

Please cite as:
Charalambous AP, Pye A, Yeung WK, Leroi I, Neil M, Thodi C, Dawes P
Tools for App- and Web-Based Self-Testing of Cognitive Impairment: Systematic Search and Evaluation
J Med Internet Res 2020;22(1):e14551
URL: https://www.jmir.org/2020/1/e14551
doi:10.2196/14551
PMID:31951218

Original Paper

Efficacy of an Electronic Health Management Program for Patients With Cardiovascular Risk: Randomized Controlled Trial

Young Ho Yun[1], MD, PhD; EunKyo Kang[1], MD; Young Min Cho[2], MD, PhD; Sang Min Park[1], MD, PhD; Yong-Jin Kim[2], MD, PhD; Hae-Young Lee[2], MD, PhD; Kyae Hyung Kim[1], MD, PhD; Kiheon Lee[3], MD; Hye Yeon Koo[3], MD; Soojeong Kim[1], BSc; YeEun Rhee[1], BA; Jihye Lee[1], MSc; Jeong Hee Min[1], BA; Jin-Ah Sim[4], PhD

[1]Seoul National University College of Medicine, Department of Family Medicine, Seoul, Republic of Korea

[2]Seoul National University College of Medicine, Department of Internal Medicine, Seoul, Republic of Korea

[3]Bundang Seoul National University Hospital, Department of Family Medicine, Seongnam, Republic of Korea

[4]Seoul National University College of Medicine, Cancer Research Institute, Seoul, Republic of Korea

Corresponding Author:
Young Ho Yun, MD, PhD
Seoul National University College of Medicine
Department of Family Medicine
103 Daehak-Ro
Jongno-Gu
Seoul, 03080
Republic of Korea
Phone: 82 27408417
Email: lawyun@snu.ac.kr

Abstract

Background: In addition to medication, health behavior management is crucial in patients with multiple risks of cardiovascular mortality.

Objective: This study aimed to examine the efficacy of a 3-month Smart Management Strategy for Health–based electronic program (Smart Healthing).

Methods: A 2-arm randomized controlled trial was conducted to assess the efficacy of Smart Healthing in 106 patients with at least one indicator of poor disease control and who had hypertension, diabetes, or hypercholesterolemia. The intervention group (n=53) took part in the electronic program, which was available in the form of a mobile app and a Web-based PC application. The program covered 4 areas: self-assessment, self-planning, self-learning, and self-monitoring by automatic feedback. The control group (n=53) received basic educational material concerning disease control. The primary outcome was the percentage of participants who achieved their clinical indicator goal after 12 weeks into the program: glycated hemoglobin (HbA_{1c}) <7.0%, systolic blood pressure (SBP) <140 mmHg, or low-density lipoprotein cholesterol <130 mg/dL.

Results: The intervention group showed a significantly higher success rate (in comparison with the control group) for achieving each of 3 clinical indicators at the targeted goal levels (P<.05). Only the patients with hypertension showed a significant improvement in SBP from the baseline as compared with the control group (72.7% vs 35.7%; P<.05). There was a significant reduction in HbA_{1c} in the intervention group compared with the control group (difference=0.54%; P≤.05). In the intervention group, 20% of patients with diabetes exhibited a ≥1% decrease in HbA_{1c} (vs 0% among controls; P≤.05).

Conclusions: A short-term self-management strategy-based electronic program intervention may improve clinical outcomes among patients with cardiovascular risks.

Trial Registration: ClinicalTrials.gov NCT03294044; https://clinicaltrials.gov/ct2/show/NCT03294044

(J Med Internet Res 2020;22(1):e15057) doi:10.2196/15057

KEYWORDS

health; hypertension; diabetes; hypercholesterolemia; randomized controlled trial

Introduction

Background

Hypertension, diabetes, and hypercholesterolemia are the global leading risks of cardiovascular mortality [1-3]. Health behaviors such as engaging in exercise, balanced diet, and weight control reduce one's risk of cardiovascular mortality. Therefore, in addition to medication, the management of health behavior is crucial in patients with multiple risks of cardiovascular mortality [4]. Clinical guidelines recommend a combined self-management strategy of health behaviors and appropriate medication use [3,5] A recent self-management approach in line with the Chronic Care Model (CCM) specifies that health behavior management should be used to manage coexisting illnesses [6-8]. Owing to the importance of self-management in patient-centered health care in combination with the increased use of mobile devices (including smartphones and tablets), there is a need to develop an efficient, affordable, and sustainable self-management strategy-based elecronig program that targets high-risk individuals [9-11].

Research concerning mobile health (mHealth) innovations to support populations with chronic illnesses and improve their health behaviors is growing [4,9,11]. A systematic review showed that the use of apps in mHealth has the potential to improve health outcomes among patients with chronic diseases through enhanced self-management [12]. A number of randomized controlled trials (RCTs) have assessed the effectiveness of mobile phone- or tablet-assisted self-management programs in addressing cardiovascular disease [13] or chronic hepatitis [14]. Although there is a need to organize intervention programs to improve the health outcomes of patients with chronic illnesses [6], few mHealth trials have addressed this [3,5].

Objectives

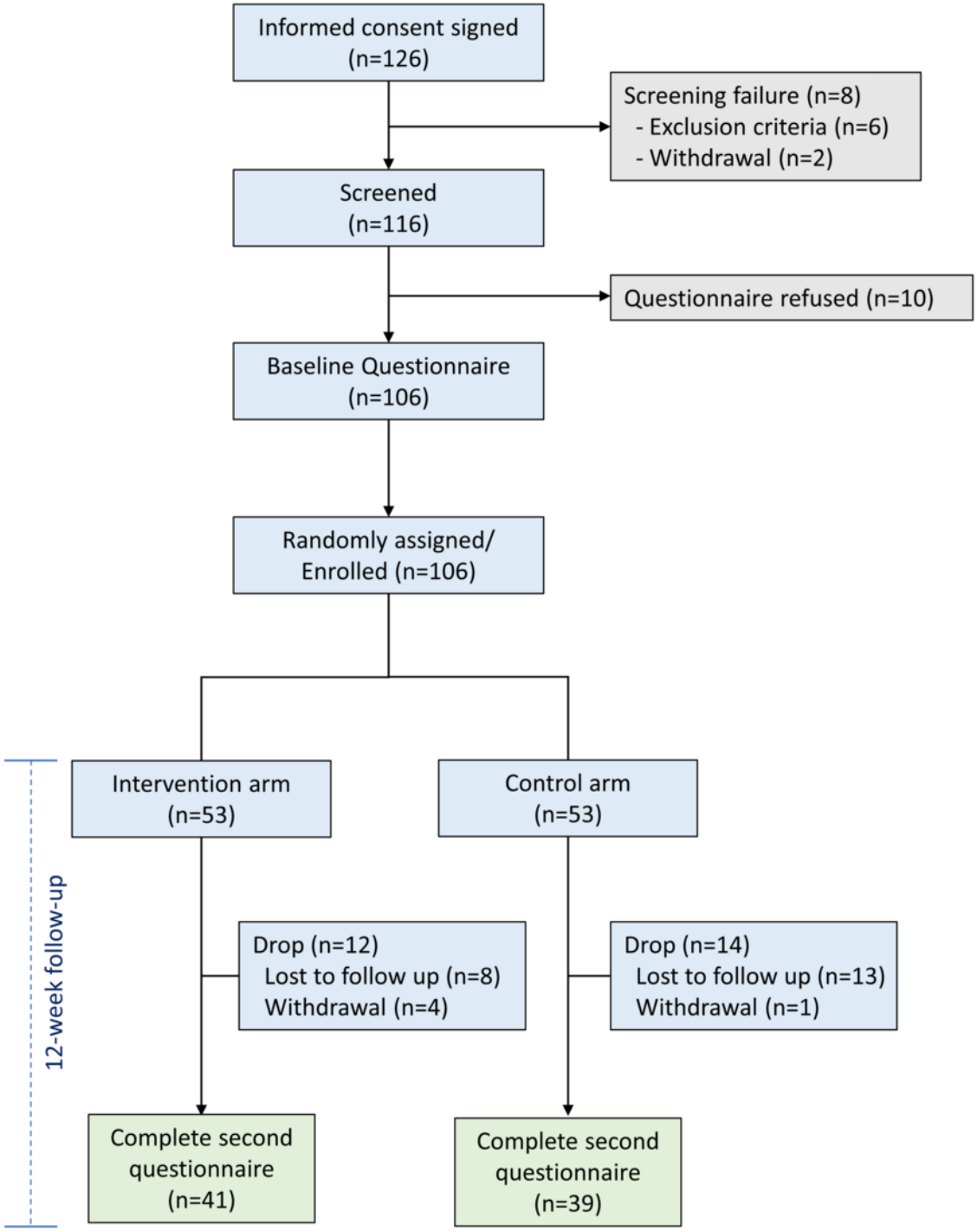
We therefore aimed to determine the efficacy of a self-management strategy-based electronic program for patients who had been treated for hypertension, diabetes, or hypercholesterolemia and who had at least one indicator of poor disease control. To do so, we provided patients with an intervention program via a Web-based health management program (mobile app or PC-Web-based) [15].

Methods

Study Design

We conducted this study with 106 patients within 2 months of treatment termination, and the patients were randomly assigned to either the control group or the intervention group (ie, *Smart Healthing*; Multimedia Appendices 1 and 2). Each physician from 2 study hospitals screened patients for the eligibility criteria by reviewing their medical records and blood test results at outpatient clinics. A clinical research coordinator at each hospital explained the study details to the participants who met the eligibility criteria (Figure 1). The patients who were eligible to participate were recruited by the physician in-charge and were asked to provide written informed consent to the researchers. The institutional review boards at the 2 hospitals approved the study protocol (numbers 1707-084-870 and B-1802/453-401). The trial was performed in accordance with the Good Clinical Practice guidelines and the Declaration of Helsinki. All staff who were involved in screening and recruiting participants were certified by their institutions for ethical conduct of research (Collaborative Institutional Training Initiatives).

Figure 1. Flowchart depicting the study methodology



Participants

From November 2017 to March 2018, we identified patients with at least one indicator of poor disease control among patients who had been treated for hypertension, diabetes, or hypercholesterolemia. We recruited patients who met the following criteria: (1) aged ≥19 years; (2) diagnosed with hypertension, diabetes, or hypercholesterolemia; (3) failed to meet 1 or more of the following clinical goals: (i) glycated hemoglobin (HbA_{1c}) <7.0%, (ii) systolic blood pressure (SBP) <140 mmHg, or (iii) low-density lipoprotein (LDL) cholesterol <130 mg/dL; (4) had a smartphone and personal computer (for the electronic program-based health care program); and (5) understood the study's purpose.

Patients were excluded from the study if they met any of the following criteria: (1) had medical conditions that would limit participation adherence (as confirmed by their referring

physician [eg, dyspnea and severe depression]); (2) could not speak, understand, or write Korean; or (3) could not understand the content of the provided materials owing to poor eyesight and/or hearing.

Randomization

We used an internet-based Clinical Research and Trial management system by the Centers for Disease Control and Prevention for participant randomization. The patients were randomly assigned (1:1) to the intervention or control group based on a random computer-generated number. To minimize the effects of potential confounding variables, we randomized participants stratified by disease type with the clinical indicators (hypertension, diabetes, or hypercholesterolemia). The research assistants executed face-to-face procedures and therefore could not be blinded when assigning participants to groups.

Control

The attention control group was encouraged to continue their usual care and routine medications and to study a health educational booklet about chronic diseases. The booklet noted 12 healthy life habits: positive thinking, regular exercise, balanced diet, proactive living, regular checks-ups, helping others, regular religious life, quitting smoking, drinking cessation, work-life balance, living with loved ones, and taking medication.

Intervention

The intervention group received the self-management strategy-based electronic program, whereas the control group received basic educational material about disease content. We developed the Smart Management Strategy for Health–based electronic program and utilized the comprehensive and multifaceted Smart Management Strategy for Health strategies. The self-management strategy-based electronic program used in this study was a 3-month Smart Management Strategy for Health Intervention, and it is comprised of an app and a Web-based program. On the basis of our literature review and interviews, we developed a conceptual framework for the Smart Management Strategy for Health intervention that incorporates management strategies for overcoming crises and developing healthy management strategies. The Smart Management Strategy for Health intervention includes the following 9 strategies: (1) assessment, (2) reality acceptance, (3) preparation for change, (4) decision making, (5) planning, (6) environment creation, (7) action, (8) feedback and maintenance, and (9) core strategies. All of these strategies can help patients overcome a disease crisis and develop healthy self-management skills [16,17].

The program covered 4 areas: self-assessment, self-planning, self-learning, and self-monitoring by automatic feedback. We targeted 4 priority areas for intervention—positive thinking, balanced diet, physical activity, and medication. The 20 learning sessions included *12 Rules for Highly Effective Health Behavior* and health management strategies.

The self-management strategy-based electronic program was used for 12 weeks. The patients were provided with a manual with detailed instructions on how to use the program to both increase its usage rate and decrease the dropout rate. Self-evaluations were conducted with regard to the participants' self-management competence and health practices before and after the program (excellent, moderate, and poor). In addition, the patients wrote health mission statements that included their life goals, health practice goals, obstacles, and methods to overcome them and detailed promises in relation to the self-management strategy-based electronic program. Self-learning was structured with the health management strategy and health information on 12 health behavior rules. The patients received daily health educational content from the self-management strategy-based electronic program. Every week, the patients learned 1 health behavior among the 4 essential rules, and they could selectively study the other 8 health behavior rules. By graphically displaying the participants' blood glucose levels, blood pressure, and weight to them, it was possible for the participants to track any changes.

The patients could create their own health management weekly plan for the 4 essential health behavior rules and monitor their progress and health. The weekly plan addressed dieting, vegetable and fruit consumption, physical activity, and daily medication schedule. More specifically, the weekly physical activity plan included the activity's type, length of time, intensity, and schedule. The self-management strategy-based electronic program included an automatic push function and alarms for the scheduled physical activities, medications, and assessments to remind participants of their plans. After 1 week, the patients were provided with feedback to motivate and help them plan for the following week. Through periodic monthly assessments, the program identified changes in their essential health behaviors and provided feedback on monthly changes through a comparison of their prior month's results to help patients change their behavior.

Measures

The primary outcome was the percentage of subjects that met the target clinical indicators (HbA_{1c} <7.0%, SB*P*<140 mmHg in clinic, or LDL cholesterol <130 mg/dL).

The secondary outcomes included the originally proposed clinical indicator outcomes—physical activity, depression, self-management strategies, and health behaviors after 12 weeks in the program. The patients' self-management strategies were assessed with a short form of the Smart Management Strategy for Health, which is a 3-set, 16-factor, 30-item tool (ie, core strategies, 10 items; preparation strategies, 10 items; and implementation strategies, 10 items) that assesses patients' abilities to overcome health-related crises [17]. Physical activity was measured with the modified version of the Godin Leisure-time Exercise Questionnaire, which is widely used, reliable, and valid [16]. The modified version adds average duration to the original questions of average frequency of light, moderate, and strenuous exercise per week. We evaluated depression with the Patient Heath Questionnaire-9 (PHQ-9). The participants were asked to measure their 12 health behaviors with 5 scales: (1) precontemplation, (2) contemplation, (3) preparation, (4) action, and (5) maintenance, which are all based on the transtheoretical model [18,19].

We assessed the proportion of patients with a ≥1.0% decrease in their HbA_{1c} level from the baseline, the proportion of patients with a ≥10-mmHg decrease in SBP from the baseline, and a ≥15% decrease in LDL cholesterol level. We also assessed the proportion of patients with either a decrease or no change in PHQ-9 score from the baseline, a ≥5- metabolic equivalent of task increase in physical activity level, a ≥10% increase in self-management strategy, and a ≥3 habits increase in the maintenance of the 12 health habits.

The participants completed baseline questionnaires before randomization at the clinics. After 12 weeks, we conducted follow-up assessments with the participants with regard to the primary and secondary outcomes. When patients did not complete a questionnaire item, the clinical research coordinator documented the reason.

Statistical Analysis

Providing 80% power to detect a 30% proportion difference in patients achieving disease control with a 2-tailed alpha value of less than .05, we calculated that it was necessary to have 42 patients per group. We predicted a 20% dropout rate and aimed at recruiting 53 patients in each group. A multiple imputation approach was used to impute scores for missing values for the intent-to-treat analysis. The imputed values were used for the covariates analyses but not for the descriptive statistics.

We used a Student *t* test or Pearson chi-square test to determine significant differences in the baseline characteristics between the intervention and control groups. We used an analysis of covariance to estimate between-group changes in the clinical outcome numbers with general linear modeling, adjusting for the baseline score and age. We compared the participants' changes from their baseline values with their values after 12 weeks in the program. Pearson chi-square test was used to assess between-group differences in the proportion of patients with improvement (overall, depression, and physical activity). Enhanced self-management strategies and health habits were also estimated by using a Pearson chi-square test.

We used STATA version 14.2 (STATA) for all statistical analyses. A two-sided *P* value <.05 was considered significant.

Results

Study Participants

The study team contacted 281 patients between October 27, 2017, and March 26, 2018. Of these, 124 patients were eligible, and 18 were excluded because of screening failures or refusal to participate. Finally, 53 were randomized to the intervention group and 53 to the control group (Figure 1). Except for age and residence, all baseline characteristics were similar between the 2 groups (Table 1). More specifically, compared with the control group, the intervention group was older, and they were more likely to reside in metropolitan areas (*P*=.001 and .04, respectively).

Table 1. Baseline characteristics of participants.

Characteristics	Intervention group (n=53), n (%)	Control group (n=53), n (%)
Age (years)		
20-49	8 (13)	18 (34)
50-59	17 (33)	21 (40)
60-69	26 (50)	8 (15)
≥70	2 (4)	6 (11)
Sex		
Male	31 (58)	29 (55)
Female	22 (42)	24 (45)
Marital status		
Married	47 (89)	45 (85)
Unmarried	4 (8)	6 (11)
Separated/bereaved	2 (4)	2 (4)
Educational status		
High school or less	18 (34)	18 (34)
≥College or university	35 (66)	35 (66)
Presence of religion		
Yes	36 (68)	27 (51)
No	17 (32)	26 (49)
Residence		
Metropolitan	39 (74)	28 (52)
Urban or rural	14 (26)	25 (47)
Monthly income (1000 KRW[a]/month)		
≤3999	14 (26)	22 (42)
4000-4999	10 (19)	9 (17)
≥5000	29 (55)	22 (42)
Employment status		
Employed	36 (68)	33 (62)
Unemployed/retired	17 (32)	20 (38)
Disease[b]		
Diabetes mellitus	26 (49)	21 (40)
Dyslipidemia	23 (43)	23 (43)
Hypertension	11 (21)	14 (26)

[a]KRW: Korean Won.

[b]Some participants have been diagnosed with more than one disease.

Success Rate for Achieving Goals

Table 2 describes the percentage of patients' achieved goals. The intervention group showed a significantly higher success rate for achieving the targeted levels for each of the 3 clinical indicators after 12 weeks, and this higher success rate remained significant when stratified by starting medication with the Mantel-Haenszel method (P<.05). With regard to disease, the patients with hypertension in the intervention group showed significant improvement compared with the control group (72.7% vs 35.7%, P=.035; Mantel-Haenszel chi-square test). These results for the patients with diabetes and hypercholesterolemia were nonsignificant.

We found a significant reduction of HbA_{1c} in the intervention group compared with the control group (0.71 vs 0.22, respectively; between-group difference=–0.54, 95% CI –0.98 to –0.11; P=.014). The patients with hypertension exhibited a greater reduction in SBP in the intervention group compared with the control group; however, this result was nonsignificant

(17.5 mmHg vs 11.6 mmHg; *P*=.41). For the patients with hypercholesterolemia, both the intervention and control groups showed a reduction in LDL cholesterol, and the between-group difference was nonsignificant (23.7 mg/dL vs 25.3 mg/dL; *P*=.72).

Table 2. Differences in clinical outcomes controlling for the primary disease.

Time point	Intervention group (n=60)		Control group (n=57)		*P* value[a]	*P* value[b]
	Success, n	Change (%)	Success, n	Change (%)		
All diseases		60		37	.01	.02
Baseline	0		0			
12 weeks	36		21			
Diabetes[c]		54		38	.35	.43
Baseline	0		0			
12 weeks	14		8			
Hypertension[d]		73		36	.07	.04
Baseline	0		0			
12 weeks	8		5			
Dyslipidemia[e]		61		35	.08	.1
Baseline	0		0			
12 weeks	14		8			

[a]All reported *P* values are 2-sided, with *P*<.05 considered as statistically significant.

[b]Stratified analysis by starting medication (Mantel-Haenszel method).

[c]Intervention: n=26; control: n=21.

[d]Intervention: n=11; control: n=14.

[e]For both intervention and control: n=23.

Differences in Other Clinical Outcomes, Health Outcomes, and Self-Management Measures

Among the intervention group, 20% of patients with diabetes exhibited a ≥1% decrease in HbA_{1c} (compared with 0% in the control group; Table 3).

In the intervention group, 73% of the participants showed a decrease or no change in depressive symptoms (vs 51% in the control group). The Smart Healthing program strengthened the implementation strategy of the modified Smart Management Strategy for Health greater in the intervention group (57.5%) than in the control group (33.3%). However, the differences in the core and preparation strategies for both the intervention and control groups were nonsignificant (*P*=.53 and .30, respectively; Table 4). There were no important harms or unintended effects observed in either group.

Table 3. Differences in clinical measures.

Differences (clinical outcomes)	Intervention group (n=53), n (%)	Control group (n=53), n (%)	*P* value
≥1.0 percentage point decrease in glycated hemoglobin level (intervention: n=25; control n=19)	5 (20)	0 (0)	.04
≥10 mmHg decrease in systolic blood pressure (intervention: n=10; control: n=9)	8 (80)	5 (56)	.25
≥15% low-density lipoprotein decrease (intervention: n=15; control: n=17)	7 (47)	7 (41)	.76

Table 4. Differences in health outcomes and self-management measures.

Differences	Intervention group (n=53), n (%)	Control group (n=53), n (%)	*P* value
Health outcomes			
Decrease or no change in PHQ-9[a] score (intervention: n=41; control: n=39)	30 (73)	20 (51)	.04
≥5 metabolic equivalent of task physical activity (intervention: n=41; control: n=39)	29 (71)	32 (82)	.23
≥Increase in 3 of the 12 health habits (intervention: n=41; control: n=39)	12 (29)	11 (28)	.92
Self-management strategies			
≥10% increase in the *Core Strategy of SAT*[b] (intervention: n=41; control: n=39)	13 (32)	15 (38)	.53
≥10% increase in the *Preparation Strategy of SAT* (intervention: n=38; control: n=39)	15 (39)	20 (51)	.30
≥10% increase in the *Implementation Strategy of SAT* (intervention: n=40; control: n=33)	23 (58)	13 (33)	.03

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

[b]SAT: Smart Management Strategy for Health Assessment Tool.

Discussion

Principal Findings

This RCT indicated that this study's self-management strategy-based electronic program effectively encouraged patients with at least one indicator of poor disease control for diabetes, hypertension, or hypercholesterolemia to meet key guideline criteria (HbA_{1c}, SBP, and LDL cholesterol). The patients with hypertension showed a significant improvement in SBP from their baseline values in comparison with the control groups. There was also a significant reduction in HbA_{1c} in the intervention group compared with the control group. We are particularly encouraged by these findings, and we posit that this study's self-management strategy-based electronic program can more effectively support disease control in comparison with the usual care strategies.

The proportion of patients with controlled hypertension increased significantly more in the intervention group than in the control group. The proportion of patients with controlled diabetes and hypercholesterolemia also increased more in the intervention group than in the control group; however, these findings were nonsignificant, which could be a result of the study's small sample size. These improvements in the primary outcomes in our trial support the findings from earlier trials with same clinical indicators. Concerning the secondary outcomes, the mean change in HbA_{1c} and the proportion of patients with a significant decrease in HbA_{1c} level from their baseline values were both higher in the intervention than the control (ie, usual care) group [3,6,20-24].

There are several possible explanations for our findings. First, the intervention strategies were based on the Smart Management Strategy for Health program. The intervention significantly increased the participants' Implementation Strategy scores for self-management. It is possible that the CCM self-management program thus helps individuals to develop preferences for how to manage their own care [7,8] It is assumed that most patients want to remain independent; however, these preferences and patients' daily behaviors may change over time because of their symptoms, the treatments they undergo and their goals [8]. Patients with chronic illnesses must manage the medical and emotional strain of their health condition(s) [8,25]. The Smart Management Strategy for Health supports patients to help them overcome a disease crisis and develop health-related self-management skills [26,27] The fact that this intervention integrates self-management strategies with electronic program in the CCM highlights how mHealth can address cardiovascular risks [28].

Second, a user - centered electronic program has the potential to improve clinical indicators among those living with chronic diseases by allowing users to obtain information from the mobile- and Web-based pages at their own pace, to flexibly review material as needed [29,30] and by facilitating the management of multiple health behaviors [11]. Third, the noted self-management program can help patients by providing immediate, easy, and continual access to the intervention [4,8,15]. This electronic program intervention may thus provide a critical route to successful chronic care. From a clinical perspective, it could be valuable to link this electronic program-based program with face-to-face or telephone counseling [31-33].

Furthermore, we did not observe any significant changes concerning the examined health habits. There are 2 possible explanations for this finding. First, the intervention might not have been intensive enough to modify patients' long-held health habits. Second, the 12-week intervention period might have been too short to observe any meaningful changes in health habits [5].

Despite well-established data on chronic disease management, its uptake into routine clinical practice remains limited. Further innovation, optimization, and rigorous research in customized

mobile technology might improve health care delivery and outcomes [5,12].

Limitations

Several limitations of our study should be noted. First, our sample included 3 types of cardiovascular risks of varying severity and a small number of patients; thus, we lacked the power to determine meaningful differences. Further studies with larger sample sizes and distinct cardiovascular risks are needed to confirm the efficacy of the intervention. Second, approximately one-quarter of the patients in the intervention group did not complete the follow-up at the 12-week mark. Missing data for these patients may have resulted in an underestimation of the efficacy of the intervention program. Third, as the patients were aware of their group, the self-reported changes in depression, physical activity, self-management strategy, and health behaviors could have been influenced by that awareness and not just by the intervention itself. Fourth, there might be attrition bias. A quarter of the intervention group was lost to follow up, and many patients of this group were older, which may be associated with less use of newer technology. This loss may lead to an overestimate of effect. Fifth, although the attention control group was encouraged to continue their usual care and routine medications and to study a health educational booklet about chronic diseases, the Hawthorne effect may be still relevant. Finally, our trial was relatively short, and we do not know whether the changes associated with the program would be maintained over a longer period. Additional research on the long-term efficacy of this intervention, including a full-scale RCT, is warranted to confirm the efficacy of this program.

Conclusions

A short-term self-management strategy-based electronic program intervention may improve clinical outcomes among patients with cardiovascular risks. More research with context-specific trials is needed to enhance these findings, to ensure the long-term generalizability and sustainability of the program, and to indicate the cost-effectiveness of this intervention.

Acknowledgments

This study was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute and by the Ministry of Health and Welfare, Republic of Korea (grant number: HI16C0455). The authors would like to thank Editage for English language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of the app.
[PNG File , 66 KB - jmir_v22i1e15057_app1.png]

Multimedia Appendix 2
Screenshot of the intervention (Smart Healthing).
[PNG File , 49 KB - jmir_v22i1e15057_app2.png]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2387 KB - jmir_v22i1e15057_app3.pdf]

References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
2. Rodgers A, Ezzati M, Vander Hoorn S, Lopez AD, Lin R, Murray CJL, Comparative Risk Assessment Collaborating Group. Distribution of major health risks: findings from the Global Burden of Disease study. PLoS Med 2004 Oct;1(1):e27 [FREE Full text] [doi: 10.1371/journal.pmed.0010027] [Medline: 15526049]
3. McManus RJ, Mant J, Franssen M, Nickless A, Schwartz C, Hodgkinson J, TASMINH4 investigators. Efficacy of self-monitored blood pressure, with or without telemonitoring, for titration of antihypertensive medication (TASMINH4): an unmasked randomised controlled trial. Lancet 2018 Mar 10;391(10124):949-959 [FREE Full text] [doi: 10.1016/S0140-6736(18)30309-X] [Medline: 29499873]
4. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. J Am Med Assoc 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]

5. Tian M, Ajay VS, Dunzhu D, Hameed SS, Li X, Liu Z, et al. A cluster-randomized, controlled trial of a simplified multifaceted management program for individuals at high cardiovascular risk (simcard trial) in rural Tibet, China, and Haryana, India. Circulation 2015 Sep 1;132(9):815-824 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.015373] [Medline: 26187183]
6. Katon WJ, Lin EH, von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010 Dec 30;363(27):2611-2620 [FREE Full text] [doi: 10.1056/NEJMoa1003955] [Medline: 21190455]
7. Unützer J, Katon W, Callahan CM, Williams JW, Hunkeler E, Harpole L, IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. J Am Med Assoc 2002 Dec 11;288(22):2836-2845. [doi: 10.1001/jama.288.22.2836] [Medline: 12472325]
8. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 2011;61(1):50-62 [FREE Full text] [doi: 10.3322/caac.20093] [Medline: 21205833]
9. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
10. He J, Irazola V, Mills KT, Poggio R, Beratarrechea A, Dolan J, HCPIA Investigators. Effect of a community health worker-led multicomponent intervention on blood pressure control in low-income patients in Argentina: a randomized clinical trial. J Am Med Assoc 2017 Sep 19;318(11):1016-1025 [FREE Full text] [doi: 10.1001/jama.2017.11358] [Medline: 28975305]
11. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
12. 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]
13. Zhang H, Jiang Y, Nguyen HD, Poo DC, Wang W. The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: a pilot randomised controlled trial. Health Qual Life Outcomes 2017 Mar 14;15(1):49 [FREE Full text] [doi: 10.1186/s12955-017-0623-y] [Medline: 28288636]
14. Jeon JH. Evaluation of a smartphone application for self-care performance of patients with chronic hepatitis B: a randomized controlled trial. Appl Nurs Res 2016 Nov;32:182-189. [doi: 10.1016/j.apnr.2016.07.011] [Medline: 27969026]
15. Appel LJ, Clark JM, Yeh H, Wang N, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968 [FREE Full text] [doi: 10.1056/NEJMoa1108660] [Medline: 22085317]
16. Yun YH, Jung JY, Sim JA, Choi H, Lee JM, Noh D, et al. Patient-reported assessment of self-management strategies of health in cancer patients: development and validation of the Smart Management Strategy for Health Assessment Tool (SAT). Psychooncology 2015 Dec;24(12):1723-1730. [doi: 10.1002/pon.3839] [Medline: 26014043]
17. Yun YH, Jung JY, Sim JA, Lee J, Noh D, Han W, et al. Development and validation of the smart management strategy for health assessment tool-short form (SAT-SF) in cancer survivors. Qual Life Res 2018 Feb;27(2):347-354. [doi: 10.1007/s11136-017-1723-1] [Medline: 29086167]
18. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
19. Prochaska JO, DiClemente CC. Toward a comprehensive model of change. In: Miller WR, Heather N, editors. Treating Addictive Behaviors. Berlin: Springer; 1986:3-27.
20. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: a proof-of-concept randomized controlled trial. JMIR Res Protoc 2013 Sep 4;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
21. Davidson TM, McGillicuddy J, Mueller M, Brunner-Jackson B, Favella A, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015 Nov 17;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
22. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved blood pressure control using an interactive mobile phone support system. J Clin Hypertens (Greenwich) 2016 Feb;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
23. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]

24. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
25. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/BF00989485]
26. Poppe L, de Bourdeaudhuij I, Verloigne M, Shadid S, van Cauwenberg J, Compernolle S, et al. Efficacy of a self-regulation-based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019 Aug 2;21(8):e13363 [FREE Full text] [doi: 10.2196/13363] [Medline: 31376274]
27. Zuidema R, van Dulmen S, Nijhuis-van der Sanden M, Meek I, van den Ende C, Fransen J, et al. Efficacy of a web-based self-management enhancing program for patients with rheumatoid arthritis: explorative randomized controlled trial. J Med Internet Res 2019 Apr 30;21(4):e12463 [FREE Full text] [doi: 10.2196/12463] [Medline: 31038461]
28. 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]
29. Green MJ, Peterson SK, Baker MW, Harper GR, Friedman LC, Rubinstein WS, et al. Effect of a computer-based decision aid on knowledge, perceptions, and intentions about genetic testing for breast cancer susceptibility: a randomized controlled trial. J Am Med Assoc 2004 Jul 28;292(4):442-452 [FREE Full text] [doi: 10.1001/jama.292.4.442] [Medline: 15280342]
30. Chambers SK, Ritterband LM, Thorndike F, Nielsen L, Aitken JF, Clutton S, et al. Web-delivered cognitive behavioral therapy for distressed cancer patients: randomized controlled trial. J Med Internet Res 2018 Jan 31;20(1):e42 [FREE Full text] [doi: 10.2196/jmir.8850] [Medline: 29386173]
31. Gielissen MF, Verhagen S, Witjes F, Bleijenberg G. Effects of cognitive behavior therapy in severely fatigued disease-free cancer patients compared with patients waiting for cognitive behavior therapy: a randomized controlled trial. J Clin Oncol 2006 Oct 20;24(30):4882-4887. [doi: 10.1200/JCO.2006.06.8270] [Medline: 17050873]
32. Beurskens AJ, Bültmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med 2000 May;57(5):353-357 [FREE Full text] [doi: 10.1136/oem.57.5.353] [Medline: 10769302]
33. Yun YH, Lee KS, Kim Y, Park SY, Lee ES, Noh D, et al. Web-based tailored education program for disease-free cancer survivors with cancer-related fatigue: a randomized controlled trial. J Clin Oncol 2012 Apr 20;30(12):1296-1303. [doi: 10.1200/JCO.2011.37.2979] [Medline: 22412149]

Abbreviations

CCM: chronic care model
HbA_{1c}: glycated hemoglobin
LDL: low-density lipoprotein
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
SBP: systolic blood pressure

Edited by M Tilney; submitted 16.06.19; peer-reviewed by HA Park, M Levine, I Mircheva; comments to author 10.08.19; revised version received 16.08.19; accepted 26.10.19; published 22.01.20.

Please cite as:
Yun YH, Kang E, Cho YM, Park SM, Kim YJ, Lee HY, Kim KH, Lee K, Koo HY, Kim S, Rhee Y, Lee J, Min JH, Sim JA
Efficacy of an Electronic Health Management Program for Patients With Cardiovascular Risk: Randomized Controlled Trial
J Med Internet Res 2020;22(1):e15057
URL: https://www.jmir.org/2020/1/e15057
doi:10.2196/15057
PMID:32012053

Original Paper

Digital Competencies and Attitudes Toward Digital Adherence Solutions Among Elderly Patients Treated With Novel Anticoagulants: Qualitative Study

Maximilian Herrmann[1,2*], MSc; Philip Boehme[1,2*], MD; Arne Hansen[1,3], MSc; Katharina Jansson[2], DVM; Patrick Rebacz[1], Dr med; Jan P Ehlers[1], MA, DVM, PhD; Thomas Mondritzki[1,2], Dr rer medic; Hubert Truebel[1,2], MD, PhD

[1]Didactics and Educational Research in Health Science, Faculty of Health, Witten/Herdecke University, Witten, Germany

[2]Research & Development, Pharmaceuticals, Bayer Aktiengesellschaft, Wuppertal, Germany

[3]Production Planning & Logistics, Johnson & Johnson Medical Gesellschaft mit beschränkter Haftung, Norderstedt, Germany

[*]these authors contributed equally

Corresponding Author:
Hubert Truebel, MD, PhD
Research & Development, Pharmaceuticals
Bayer Aktiengesellschaft
Apratherweg 18
Wuppertal, 42113
Germany
Phone: 49 202360
Fax: 49 202364115
Email: hubert.truebel@bayer.com

Abstract

Background: Nonadherence to medication is a driver of morbidity and mortality, and complex medication regimens in patients with chronic diseases foster the problem. Digital technology might help, but despite numerous solutions being developed, none are currently widely used, and acceptance rates remain low, especially among the elderly.

Objective: This study aimed to better understand and operationalize how new digital solutions can be evaluated. Particularly, the goal was to identify factors that help digital approaches targeting adherence to become more widely accepted.

Methods: A qualitative study using a conceptual grounded theory approach was conducted. We included patients aged 65 years and older who routinely took new oral anticoagulants. To generate theses about the digital competencies of the target group with daily medication intake, face-to-face interviews were conducted, recorded, and anonymized. After coding the interviews, categories were generated, discussed, and combined with several theses until saturation of the statements was reached.

Results: The methodological approach led to the finding that after interviews in 20 of 77 potentially available patients, a saturation of statements was reached. The average patient's age was 75 years, and 50% (10/20) of the subjects were female. The data identified five main coding categories—Diseases and medicine, Technology, Autonomy, Patient narrative, and Attitude toward technologies—each including positive and negative subcategories. Main categories and subcategories were summarized as Adherence Radar, which can be considered as a framework to assess the potential of adherence solutions in the process of prototyping and can be applied to all adherence tools in a holistic manner.

Conclusions: The Adherence Radar can be used to increase the acceptance rate of digital solutions targeting adherence. For a patient-centric design, an app should be adapted to the individual patient's needs. According to our results, this application should be based on gender and educational background as well as the individual physician-patient relationship. If used in a proper, individualized manner, digital adherence solutions could become a new cornerstone for the treatment of chronically ill individuals.

(J Med Internet Res 2020;22(1):e13077) doi:10.2196/13077

KEYWORDS

medication adherence; eHealth; mHealth; digital health; smartphone; elderly patients; compliance; digital device; digital competencies; grounded theory; delivery of health care; diffusion of innovation

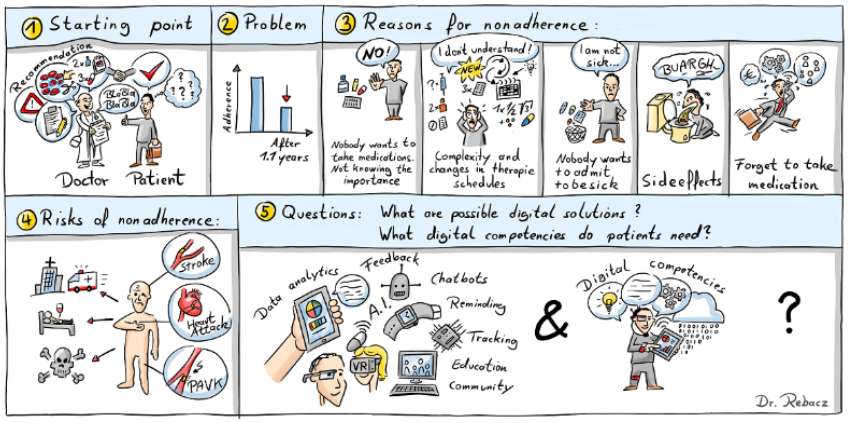
Introduction

Background

Nonadherence to prescribed medications is a driver of morbidity and mortality. A recent report of the World Health Organization (WHO) has shown that approximately 50% of the patients are nonadherent to their medication in developed countries, and that percentage is even higher in middle- and low-income countries [1,2]. Nonadherence to medication in case of chronic diseases has different reasons and leads to an increased risk for hospitalization and death (Figure 1).

Nonadherence usually worsens when patients with chronic diseases must adhere to complex medication regimens [3]. The total cost of nonadherence is estimated to be between US $100 and US $300 billion, and nonadherence is also responsible for more than 125,000 deaths per year in the United States alone [4-6]. For years, the research on chronic diseases has shown that in a setting of a clinical trial, adherence can be increased by using nondigital ("digital" is defined as continuous data collection and instant feedback) devices [7]. In a trial of 2 months, Laster et al [8] were able to show an increase in adherence of 12.7% for patients with glaucoma using an electronic medication alarm device, which displays the last time the bottle was opened. Rosen et al [9] also showed an increase in adherence of 15% for patients with diabetes using an electronic monitoring cap, which records the date and time the bottle was opened in a trial of 3.5 months. Despite the remarkable increase of adherence in clinical trials, these solutions did not translate into the treatment of patients. Reasons for that could be the inconvenient use of the readout by health care personnel, the delayed feedback, or the associated costs [7].

Figure 1. Research question: How digital competencies can influence adherence?



Objectives

To better understand how new digital devices can increase adherence in a real-life scenario and why no solution has reached broad acceptance yet, we conducted a qualitative study to (1) identify factors that make digital solutions successful and (2) generate theses to understand which digital competencies patients aged 65 years and older on anticoagulants should have, to be able to use a digital solution to increase adherence. The sample selected was a population of patients that routinely took new oral anticoagulants (NOACs). Adherence to this type of medication is especially important because NOACs need to be taken daily to prevent the occurrence of thrombotic or embolic events [10,11].

Methods

Design

To learn how to design a digital adherence solution, we generated theses about the digital competencies of a target group with daily medication intake. Therefore, a qualitative study according to Mayring [12], with a conceptual and theoretical approach based on the Straussian *grounded theory* [13-15], was conducted in patients aged 65 years and older on anticoagulants. Face-to-face guided interviews were performed to generate theses. The responses were recorded and anonymized. Afterward, the interviews were evaluated qualitatively using the software MAXQDA (Version 13, VERBI Software, Berlin, Germany) by two independent examiners. After coding of the interviews was accomplished, categories were generated, discussed, and combined with several theses until saturation of the statements was reached.

Recruitment of Participants

General practitioners associated with Witten/Herdecke University cooperated to identify participants for this study. Inclusion criteria comprised age (>65 years) and the intake of an NOAC. No further exclusion criteria or screening questionnaires were applied. Cooperating general practitioners contacted potential participants and asked them to volunteer. A cover letter clarified the intent of the study. The privacy policy was provided via an additional letter. A sampling procedure was performed in a probabilistic manner. All participants were informed about study details, for example, duration of the study interview or data storage policy. Participants willing to consent were asked to send a reply letter with their signed consent form in a prepaid envelope.

The Ethics Committee at Witten/Herdecke, University Faculty of Medicine, authorized the study and its ethical and legal implications (statement no. 89/2017).

Data Collection

Data were collected from August 2017 to December 2017. Face-to-face interviews were either conducted at the

participants' home or at Witten/Herdecke University. Each interview was audio recorded with the participant's permission. The study methods and results were reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) [16]. The completed COREQ checklist can be found in Multimedia Appendix 1.

Data Analysis and Statistics

To evaluate the digital competencies of the participants, an open interview guideline was created (Multimedia Appendix 2) by an expert panel with a multiprofessional background in medicine and engineering, pharma, psychology, and economics (for more details on the research team and reflexivity, see Multimedia Appendix 1). The interview guideline was previously tested in 5 randomized participants (who were subsequently not included in the results) in pilot interviews and was reviewed and revised by the expert panel. The process was supported by a literature review of recent publications on digital competencies and usability of various types of technical and digital solutions to increase adherence. These types of technical and digital solutions range from interactive notification apps, which remind the patient to take their medication, and education-based apps (eg, Transplant Hero and Incendant 360° Patient Education Suite) to smart wireless pill bottles, which alert the patient and send a message to the patient's phone (eg, Smart Pill Bottle of AdhereTech). In addition, solutions that use an ingestible sensor to measure the intake of the medication in stomach (eg, Proteus Discover) or artificial intelligence technologies that leverage a visual recognition algorithm to monitor patient adherence (eg, AiCure) were evaluated. Finally, we assessed an intelligent and socially interactive health care robot intended to help patients with their disease management (eg, Mabu Personal Healthcare Companion). To generate theses, the conceptual approach of *Grounded Theory*, a method applicable when investigating social processes related to complex phenomena, was used (for more details on the study design, see Multimedia Appendix 1). The approach is based on the subjective experience of participants. Straussian *grounded theory* was used to analyze the generated data in the following three stages: open coding, axial coding, and selective coding. A special focus was placed on the open coding approach of the *grounded theory* method. This approach needed to be confirmed in additional theory formation and further studies. Memos were written continuously to document the conceptual and theoretical ideas that emerged when exploring the data [13-15]. Audio-recorded interviews were coded directly by two examiners without transcription by using software MAXQDA (VERBI Software; for more details on analysis and findings, see Multimedia Appendix 1). Table 1 shows 2 examples of the coding process.

The comparisons between the groups for category parameters in the quantitative and social demographic parts of the questions were achieved with the Fisher exact test. The difference was defined as statistically significant if a value of $P<.05$ was reached.

Table 1. Examples for creating categories.

Quote	Open coding	Axial coding (subcategory)	Selective coding (core category)
"The more I know about the disease, the more crazy obsessed I become."	Knowledge, disease, more crazy obsessed	Increasing knowledge of a medical condition can be worrisome for the patient	Diseases and medicine
"I don't want to know so much about my illness, because then I just worry too much about it."	Knowledge, illness, worry too much	Increasing knowledge of a medical condition can be worrisome for the patient	Diseases and medicine
"From my point of view digital solutions are important for life-threatening conditions where I need to take a medication."	Digital solution, life-threatening condition, medication	Digital solutions should be developed for life-threating illness	Diseases and medicine
"I do not want a medical app unless it is essential for life and absolutely necessary for my illness."	Medical app, essential for life, necessary for illness	Digital solutions should be developed for life-threating illness	Diseases and medicine
"Digitisation can't be stopped, but you have to be careful not to lose your independence if you rely too much on digital solutions."	Digitization, loss of independence, reliability	Too much reliance on digital solutions can lead to loss of independence	Autonomy
"By digital solutions and a control of the intake of my medication, I would feel incapacitated."	Digital solution, control, feel incapacitated	Too much reliance on digital solutions can lead to loss of independence	Autonomy
"Nowadays I find the constant availability due to digital solutions annoying."	Constant availability, digital solutions, annoying	Digital solutions can promote the feeling of being surveilled	Autonomy
"I don't want to be a puppet of digital solutions which want to control me constantly."	To be a puppet, digital solutions, constant control	Digital solutions can promote the feeling of being surveilled	Autonomy

Results

Participants

A total of 77 participants were identified at family practitioners' offices affiliated with Witten/Herdecke University. We subsequently recruited patients and performed interviews until saturation of the statements was reached (after N=20; Table 2). Participants were defined as *nonadherent* if they did not take their prescribed medication once in the last 4 weeks. Otherwise, participants were classified as *adherent*. The classification of nonadherence was based on participant self-report of their medication adherence. The average age was 75 years, and 50% (10/20) of the patients were female. Moreover, 60% (12/20) of the participants were married or lived in a partnership, 30% (6/20) were single or widowed, and 10% (2/20) lived in a more generational household. Of the 20 participants, 11 (55%) had from atrial fibrillation. In addition, 75% (15/20) took Eliquis (apixaban), 20% (4/20) took Xarelto (rivaroxaban), and 5% (1/20) took Pradaxa (dabigatran). At the time of the study, the average duration of NOAC usage was 29.3 months. Of the 20 participants, 60% (n=12) used a smartphone, 20% (n=4) used a mobile phone, 15% (n=3) used a senior mobile phone (mobile phone especially for seniors with, for example, large buttons and large display), and only 5% (n=1) had no phone. Of the 20 participants, 13 (65%) used other digital devices than a phone. In total, 60% (12/20) reported being adherent.

Table 2. Participant demographics.

Patient	Sex	Age (years)	Profession	Marital status	Indication	Medication	Medication before NOAC[a]	Duration on NOAC (month)	Type of phone	Other digital devices	Do you forget to take your medication?
1	M[b]	73	Mechanical engineer	S/W[c]	ST[d], CA[e]	Xarelto	ASA[f]	36	SP[g]	—[h]	Yes
2	F[i]	83	Qualified salesperson	S/W	ST	Eliquis	Marcumar	36	MP[j]	—	No
3	F	69	Personnel administrator	S/W	Afib[k], PM[l], TB[m], ST	Pradaxa	Marcumar	144	MP	PC[n]	No
4	M	72	Diploma in public administration	M/P[o]	Afib	Eliquis	Marcumar	6	SP	PC, Tablet	Yes
5	M	74	Engineer	MGH[p]	CA	Eliquis	—	24	SP	PC	No
6	M	68	IT sales staff	M/P	ST	Xarelto	—	3	SP	Tablet, PC	No
7	F	70	Hairdresser	M/P	CA, PM[q]	Eliquis	Marcumar	5	MP	—	No
8	F	76	Secretary	S/W	TB	Eliquis	Xarelto	60	SP	PC	No
9	F	88	Housewife	S/W	Afib	Eliquis	—	7	NMP[r]	—	Yes
10	F	72	Teacher	M/P	CA, valve does not work properly	Xarelto	Marcumar, Eliquis	3	SP	—	No
11	F	82	Tailor	MGH	TB	Eliquis	—	12	SMP[s]	Tablet	No
12	M	77	Postal service employee	M/P	Afib	Eliquis	ASA	4	SMP	—	Yes
13	M	68	Teacher	M/P	Afib	Eliquis	—	24	SP	Laptop	Yes
14	M	85	Electrical engineer	S/W	Afib, PM	Xarelto	Marcumar	36	SP	PC, laptop, and tracker	Yes
15	M	70	Pharma sales representative	M/P	Afib	Eliquis	—	17	SP	PC	No
16	M	72	Lawyer	M/P	Afib	Eliquis	—	24	SP	PC	Yes
17	F	76	Pharmacist	M/P	Afib	Eliquis	—	24	SP	PC, tablet, and heart rate watch	No
18	M	75	Businessman, reporter, publisher	M/P	Afib	Eliquis	—	108	SP	PC, Apple watch, and tablet	Yes
19	F	78	Industrial management assistant	M/P	Afib	Eliquis	—	9	MP	—	No
20	F	78	Childminder	M/P	Afib, CA	Eliquis	—	4	SMP	Tracker and tablet	No

[a]NOAC: new oral anticoagulant.

[b]M: male.

[c]S/W: single/widowed.

[d]ST: stroke.

[e]CA: cardiac arrhythmia.

[f]ASA: acetylsalicylic acid.

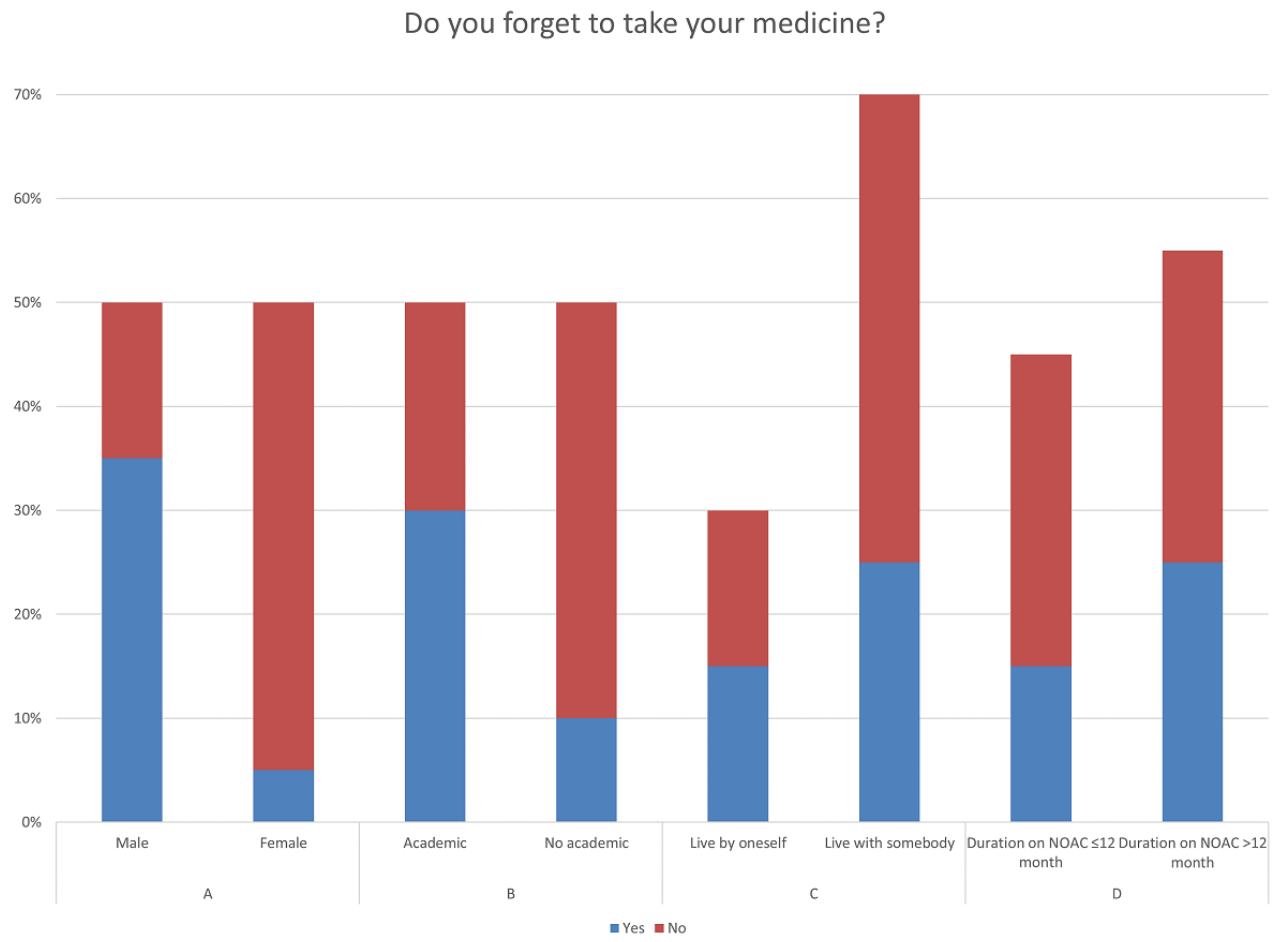
[g]SP: smartphone.

[h]No medication before NOAC/no other digital devices.

[i]F: female.
[j]MP: mobile phone.
[k]Afib: atrial fibrillation.
[l]PM: pacemaker.
[m]TB: thrombosis.
[n]PC: personal computer.
[o]M/P: married/partnership.
[p]MGH: more generational household.
[q]PM: pacemaker.
[r]NMP: no mobile phone.
[s]SMP: senior mobile phone.

As shown in Figure 2, more female participants (90%; 9/10) reported being adherent in contrast to male participants (30%; 3/10). The Fisher exact test indicated statistical evidence that female participants were significantly more likely to report being adherent than male participants (Fisher exact test, P_{2A}=.02). The results also show that more academics (6/10, 60%) reported nonadherence than nonacademics (Figure 2). Although not statistically significant (Fisher exact test, P_{2B}=.67 and P_{2C}=.07), more participants (50%; 3/6) who lived by themselves reported being nonadherent than participants (36%; 5/14) who lived in company (Figure 2). Of participants taking NOACs ≤12 months, only 33% (3/9) reported nonadherence, and longer medication duration led to an increase in reporting of nonadherence (5/11, 46%; Fisher exact test, P_{2D}=.64), as shown in Figure 2.

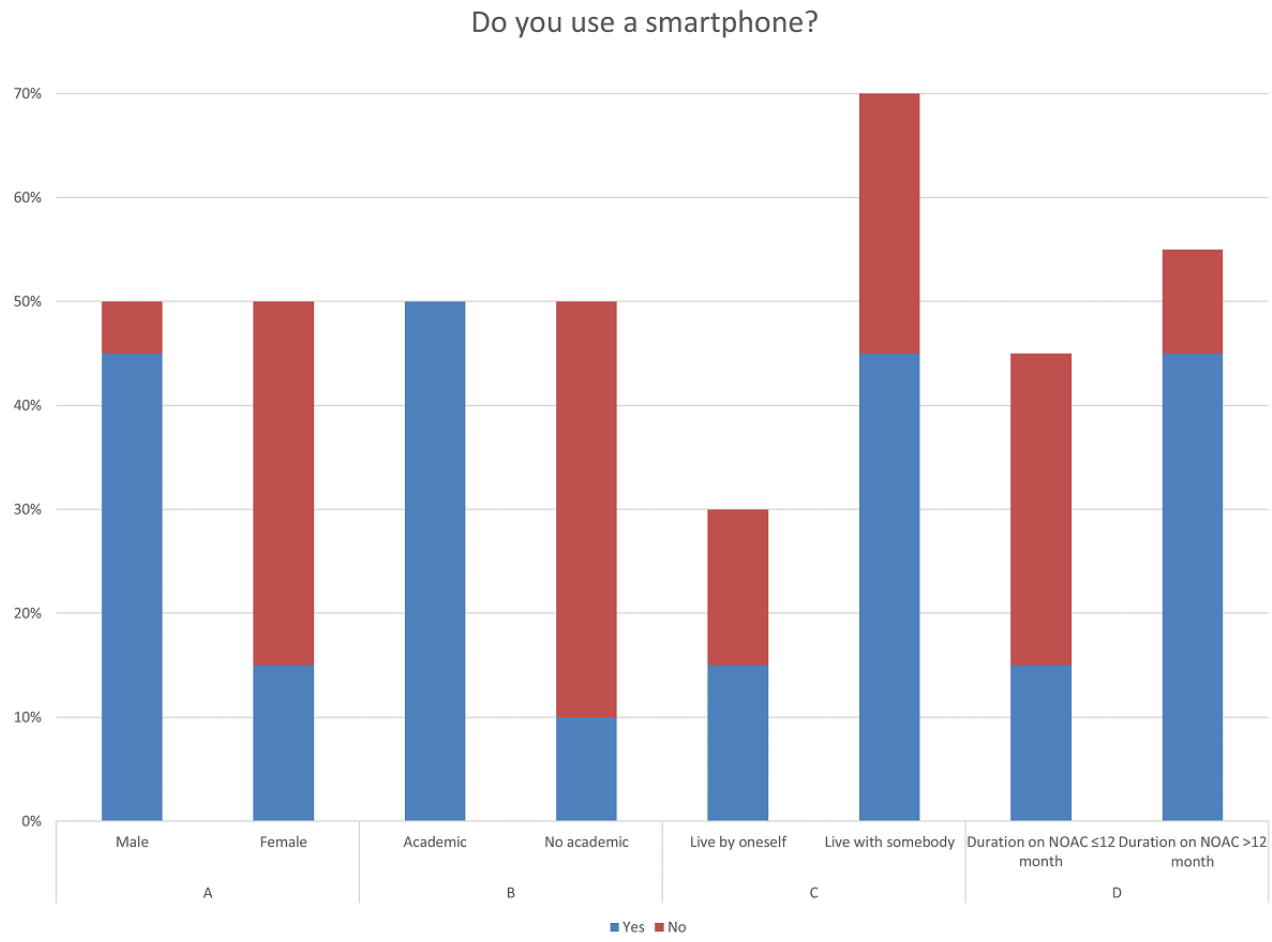
Figure 2. Participants-reported adherence by attributes.



As shown in Figure 3, most male participants (9/10, 90%) used a smartphone, whereas fewer female participants reported using a smartphone (3/10, 30%). In Figure 3, all academics (N=10) used a smartphone in contrast to nonacademics (2/10, 20%). There is statistical evidence that male participants were significantly more likely to use a smartphone than female participants (Fisher exact test, P_{3A}=.02) and that academics were significantly more likely to use a smartphone than nonacademics (Fisher exact test, P_{3B}<.001). As shown in Figure 3, fewer participants (3/6, 50%) who lived by themselves used a smartphone as compared to the participants (9/14, 64%) who lived with somebody, and fewer participants (3/9, 33%) who

were taking NOACs for ≤12 months used a smartphone in comparison with participants (9/11, 82%) who were taking NOACs >12 months. The last two results in Figure 3 were not statistically significant (Fisher exact test, P_{3C}=.64; Fisher exact test, P_{3D}=.67).

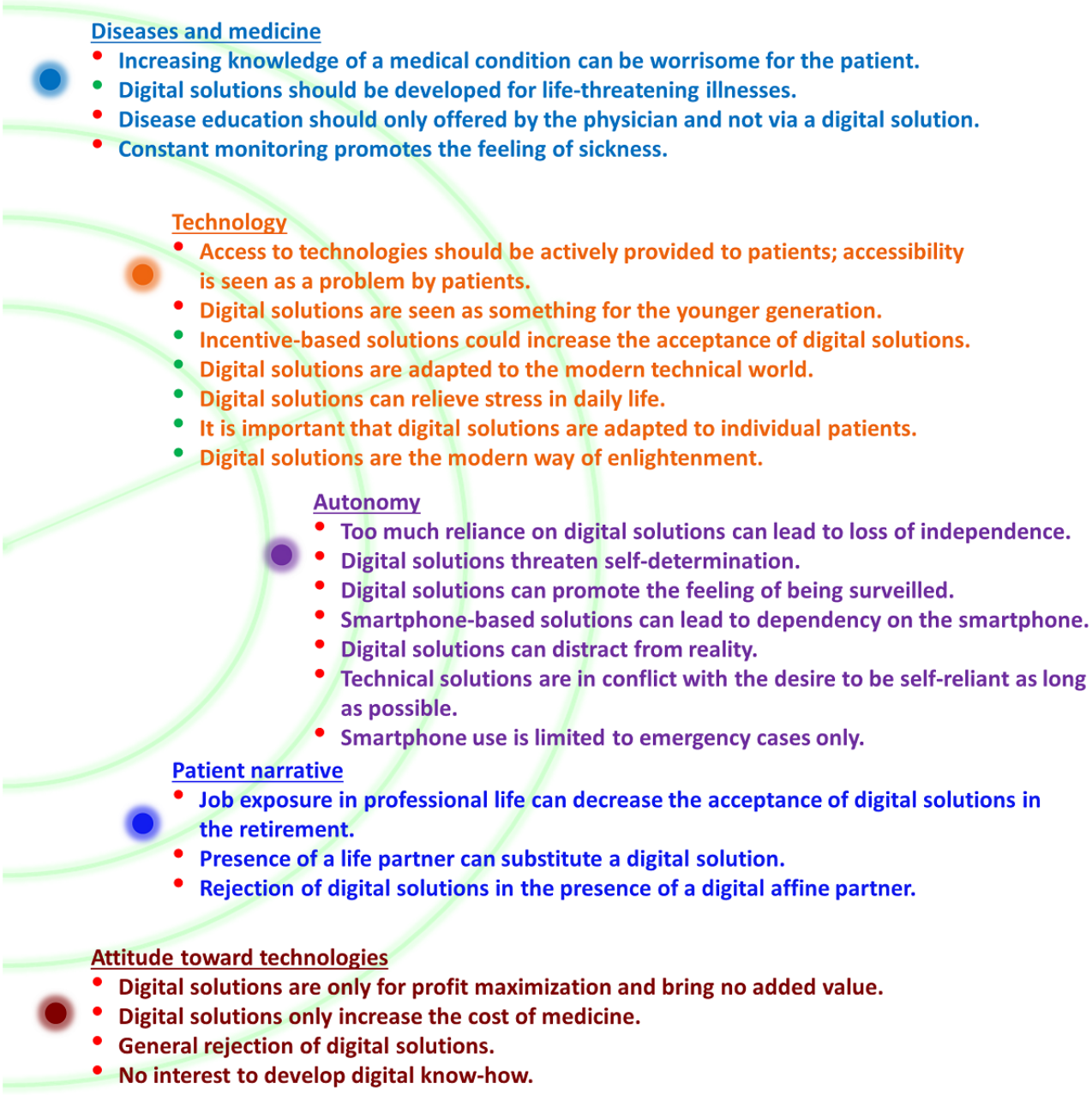
Figure 3. Participant smartphone use by attributes.



Open Interview Results and Development of an Adherence Radar

In open interviews, participants were asked about their general use of digital devices, especially smartphones, and their experience with these digital devices. Moreover, patients were shown different digital solutions to increase adherence. The results of the Straussian *grounded theory* approach, where we used three stages—open coding, axial coding, and selective coding—to analyze the generated data, could identify the following five main categories, and each main category includes positive and negative subcategories (Figure 4).

Figure 4. Adherence Radar: Main and subcategories. Green bullet points denote positive and red bullet points denote negative subcategories.



1. Diseases and Medicine: This category deals with the general feelings and thoughts of the patients toward diseases, for example, “The more I know about the disease, the more crazy obsessed I become.” Furthermore, this main category is about the knowledge of the patient about medicine, for example, “I feel fully informed by my physician about my medicine and do not need digital support.”
2. Technology: This category addresses, especially, the assessment and technological acceptance of the patients toward digital solutions in the health care sector, for example, “Digital solutions in the health care sector are good, they are adapted to today's technical world” as well as barriers of these technologies in a daily life, for example, “In general, I am interested in new technologies but have no access.”
3. Autonomy: This category deals with the general feeling of dependency toward the digital transformation of the health care sector, for example, “Digitization is unstoppable, but you have to be careful not to lose your independence,” and with the freedom from external control or influence of digital solutions in the health care sector, for example, “I do not want to become the puppet of systems.”
4. Patient narrative: This category summarizes the general thoughts toward the use of digital solutions in the health care sector, for example, “I've had enough to do in my professional life with digital things, in my retirement I have no interest in it anymore,” and includes the willingness and interest of a digital solution when they live in a partnership, for example, “I do not need digital solutions, my partner supports me in taking my medication regularly.”
5. Attitude toward technologies: This category covers the attitude toward using a digital solution, for example, “I am no longer interested in the digital age and do not want to accept it anymore,” and the risk of misuse of such

technologies in the health care sector, for example, “Digital solutions are of no use to me, they only increase the cost of medicines.”

Discussion

Adherence

Here, we report on a population of patients aged 65 years and older treated with NOACs. These patients were chosen because regular intake is required for sufficient protection from adverse events such as stroke or embolism. A *grounded theory* approach [13-15] was applied to a rather small sample size until saturation of statements was reached. In general, the reported adherence of our population was in line with previously published studies. For instance, the WHO reported that approximately 50% of patients, in general, are adherent [3,4], and a meta-analysis of data on 376,162 patients showed 57% of patients were adherent [17]. This is similar to our results, with 60% reported adherence as shown in Figure 2.

Surprisingly, and in contrast to the study by Manteuffel et al [18], our data indicate an influence of gender. Female participants were more likely to report being adherent than male participants. This influence needs to be considered when adherence solutions are developed.

Figure 2 shows that more academics reported being nonadherent. A reason for that could be higher rates of distrust of academics toward their physician in contrast to nonacademics and maybe a tendency to question the results of the physician [19].

The result in Figure 2 indicates that patients who live by themselves have higher risks of being nonadherent, which is in line with the results of Uchino [20] that social support provides survival advantages to patients with various diseases. Our results confirm the observation of several studies conducted before that reported a drop in adherence over time and especially after the first year of treatment (Figure 2) [21-23].

Smartphone and Adherence

In accordance with Anderson and Perrin [24] who showed an increase in smartphone use in adults aged 65 years and older from 18% in 2013 to 42% in 2016, our results (Figure 3) show a total smartphone use of 60%. In contrast to female participants, most male participants used a smartphone (90%). This is in line with our other findings that adherence solutions should be gender specific to eliminate the potential problem of a wide demographic spread of the target group (eg, men or women, young or old, and highly or less educated). The results of the study by Anderson and Perrin [24] also confirm our finding displayed in Figure 3 that there was a positive correlation between educational level and smartphone use.

These results and the observation that more academics forgot to take their medicine (Figure 2) strongly suggest the need for creating a smartphone-based adherence solution, especially for academics. It is implied that adherence solutions should always be designed specifically for the customer and individually for the patient. This should not suggest that each smartphone app has to be developed individually for each patient, nor that the software itself should have a unique code. What can be suggested is that each app should be designed to cover the most important adherence factors of different patient groups, that is, the app should include factors such as illness severity and duration, age, gender, and patient’s level of education. In this way, the app can be customized to different patient groups. However, to tailor apps in the future, it does not appear unrealistic to make use of the enormous power of artificial intelligence and machine learning. These technologies continuously analyze the patients and their behaviors while allowing them to improve themselves and adapt according to the gained insights, thereby providing the most individual support.

Furthermore, the cultural background and its impact on technology acceptance should also be taken into account. Several studies have shown the importance of the cultural background for the uptake and use of technology [25-27]. Alagöz et al [28] showed that in contrast to German participants, Polish and Turkish participants significantly increased their acceptance of medical technology as they aged. Alagöz et al [28] hypothesized that this is because of the economic gap and the difference in history between Germany and these countries; this shows that a deeper understanding of the factors underlying technology acceptance, beyond national borders and cultural contexts, is needed. Alsswey et al [29] pointed out that most elderly Arab participants accepted a mobile health apps, which was based and designed on Arab cultural background, and it is important to integrate cultural aspects as well as personal characteristics and experiences into the design process of a mobile health app [29].

If the results shown in Figure 2 are linked to those in Figure 3, it becomes clear that nonadherent patients who live by themselves are also less likely to use a smartphone. Thus, patients who are mostly on their own could be supported with personalized digital adherence solutions, which should include interpersonal relationships. Prochaska and Velicer [30] also highlighted this result using the Transtheoretical Model, developed by Prochaska and DiClemente [31]. They were able to show that individuals must go through six phases of change to change health behavior, and the most promising improvement in computer-based programs is interpersonal contact.

However, looking at the results in Figures 2 and 3, in our cohort, more participants reported being nonadherent after 12 months of NOAC use, but smartphone use among participants was also much higher. This indicates that one could achieve great success with a smartphone-based adherence solution in a long-term medical treatment.

Adherence Radar

As already shown in several literature reviews, there is a huge bandwidth of new digital adherence solutions in the health care sector with different approaches solving the problem of nonadherence [21,32]. Despite that, none have proven to be successful outside of a clinical study or the pilot phase.

To develop a *new* digital solution targeting that problem, customer feedback seems to be the key. Therefore, we developed a framework to assess the success potential of adherence solutions already in the process of prototyping. Furthermore,

the *Adherence Radar* is an analogy and could also provide orientation in a wide range of existing adherence solutions. For this, each main category (as well as related subcategories) should be used to analyze and characterize an already existing or imagined solution. To exemplify, we show how the *Adherence Radar* can be applied to all adherence tools in a holistic manner.

Adherence Solutions Including Educational Interventions

Educational interventions, as shown by Shah et al [33], can have a great impact on adherence, but as shown in the *Adherence Radar*, several issues need to be considered (category *Diseases and medicine*; Figure 4). Patients stated that they trust their physicians most, and they do not want them to be replaced regarding their disease education by a digital support tool. Furthermore, they stated that they do not necessarily want to know more about their diseases because this would lead to a kind of hypochondria (Main category *Diseases and medicine*; Figure 4). Another point is that education has to be in line with the severity of the underlying illness (Figure 4). We hypothesize that by complying with the *Adherence Radar*, a solution could offer personalized education by the treating physician in the right dose that needs to be defined by the patient.

Adherence Solutions Using Sensors to Detect Medication Intake

Sensor-based adherence solutions make use of the smartphone to confirm the intake of the medication via visual [34] or ingestible [35] sensors. Such a sensor-based solution could help significantly in severe disease conditions because the patient feels fully cared for. Besides that, these solutions are technically more sophisticated and provide wide flexibility. In the main category *Autonomy* of the *Adherence Radar*, patients stated that they do not want to lose their independence. They worry that the constant digital monitoring promotes the feeling of being sick. An additional aspect that can be addressed using the *Adherence Radar* is one of data privacy and safety. Patients do not want to be monitored in such a close manner and want to act self-determined, points that can be found in the *Adherence Radar* in the main categories *Attitude toward technologies* and *Patient narrative*. The acceptance of a suitable solution is therefore largely dependent on the factors of privacy and the feeling of autonomy.

Smartphone-Independent Adherence Solutions

The smartphone itself enables some of the adherence solutions to piggyback on a digital device that most of the participants in our study already used. Despite that, other approaches use non–smartphone-based solutions. Examples include robots and other devices as well as a digital pill bottle [36-38]. According to our *Adherence Radar*, bringing in a new device is critical, as the interviewed patients reported that they were already afraid of being dependent on a single digital device. They feared losing their *Autonomy*. It is also possible that the patient feels too much controlled by this additional device, and thus, the self-determination would be lost. In addition, it would touch the *Patient narrative* that increasing costs for digital solutions should be avoided. It needs to be stated that despite all worries, *Technology* itself was widely seen in a positive manner in our study. However, overall, from our point of view, it will be much harder to attain wide acceptance for a non–smartphone-based solution to become routine operation.

Limitations and Further Research

Our study, to identify factors that make digital solutions successful and generate theses to understand which digital competencies patients aged 65 years and older on anticoagulants should have, to be able to use a digital solution to increase adherence, has a few limitations. The results of the study can only be seen as a snapshot of the topic, which is constantly changing because of the ongoing digitalization. Furthermore, the number of participants seems to be small (N=20), which is because of the chosen approach of the *Grounded Theory*, as the saturation of the statements was reached after 20 interviews. Our study can only be seen as a first step for such a broad and important topic, and further research with a different approach, and therefore a larger number of participants, should be pursued. As further steps in the research, it is necessary to conduct a quantitative confirmation of the factors of the *Adherence Radar* with questionnaires with more than 500 seniors. Additional observational studies are needed to test the use of apps and devices in older people. This could be compared with the group of people aged 55-65 years as the soon-to-be seniors. Thereafter, the vision is to transform the insights of the *Adherence Radar* into a score to compare different adherence solutions clearly.

Conclusions

The aim of this study was to identify factors that make digital adherence solutions in a real-life scenario successful and how elderly patients can use such a solution. We learned that technology itself is not the problem of a limited uptake and negative view toward digital adherence solutions. As our *Adherence Radar* in Figure 4 demonstrates, the subcategories in the main categories, with the exception of *Technology*, are generally more negatively affected. Here, it becomes clear that digital solutions are partly seen as tools for the younger generation and as gimmickry, but in general, adults aged 65 years and older are open to new technologies as well as digital solutions, and this trend will automatically increase over time, as the aging digital-affine generations will follow. However, it is important that easy access to these new solutions is guaranteed, and these solutions are adapted to the individual patient's needs. A key element for a successful adherence solution seems to be that it is always designed in a customer-specific manner and uniquely for each patient group. Here, not only gender but also educational background seems to play a role; in addition, the physician-patient relationship is an important factor. The patient must not be made to feel like he/she is losing autonomy and controlled externally, but that he/she is actively and individually supported in his/her medication intake via a digital solution. In our opinion, the smartphone itself seems to be a suitable medium to develop an adequate digital adherence solution for patients because no additional device is needed.

In conclusion, digital adherence solutions can improve the standard of care and help reduce complications of nonadherence. We have shown that there is no universal solution, and tailor-made solutions will be needed.

Our *Adherence Radar* can be a cornerstone in the development of such a solution, for instance, in a design thinking approach, as it helps shed a different light on adherence solutions, in general, and helps people ask the right questions to the right patient.

Acknowledgments

We would like to thank the expert panel that helped to develop the open interview guideline and the patients who spent a lot of time answering the questions.

Conflicts of Interest

MH, PB, KJ, TM, and HT were full-time employees of Bayer Aktiengesellschaft, and AH was a full-time employee of Johnson & Johnson Medical Gesellschaft mit beschränkter Haftung at the time of preparation of the manuscript.

Multimedia Appendix 1
Consolidated Criteria for Reporting Qualitative Reasearch (COREQ): 32-item checklist.
[DOCX File , 36 KB - jmir_v22i1e13077_app1.docx]

Multimedia Appendix 2
Interview guideline.
[DOCX File , 30 KB - jmir_v22i1e13077_app2.docx]

References

1. World Health Organization. 2003. Adherence to Long-term Therapies: Evidence for Action URL: https://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf;jsessionid=72F3A7EFF7571CB19746095A6CE0E979?sequence=1 [accessed 2019-11-06] [WebCite Cache ID 6pIt6ezF8]
2. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
3. Laliberté F, Nelson WW, Lefebvre P, Schein JR, Rondeau-Leclaire J, Duh MS. Impact of daily dosing frequency on adherence to chronic medications among nonvalvular atrial fibrillation patients. Adv Ther 2012 Aug;29(8):675-690. [doi: 10.1007/s12325-012-0040-x] [Medline: 22898791]
4. Benjamin RM. Medication adherence: helping patients take their medicines as directed. Public Health Rep 2012;127(1):2-3 [FREE Full text] [doi: 10.1177/003335491212700102] [Medline: 22298918]
5. McCarthy R. The price you pay for the drug not taken. Bus Health 1998 Oct;16(10):27-8, 30, 32. [Medline: 10185113]
6. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
7. Boehme P, Wienand P, Herrmann M, Truebel H, Mondritzki T. New digital adherence devices could prevent millions of strokes from atrial fibrillation by the end of the next century. Med Hypotheses 2017 Oct;108:46-50. [doi: 10.1016/j.mehy.2017.07.034] [Medline: 29055399]
8. Laster SF, Martin JL, Fleming JB. The effect of a medication alarm device on patient compliance with topical pilocarpine. J Am Optom Assoc 1996 Nov;67(11):654-658. [Medline: 8979657]
9. Rosen MI, Rigsby MO, Salahi JT, Ryan CE, Cramer JA. Electronic monitoring and counseling to improve medication adherence. Behav Res Ther 2004 Apr;42(4):409-422. [doi: 10.1016/S0005-7967(03)00149-9] [Medline: 14998735]
10. Borne RT, O'Donnell C, Turakhia MP, Varosy PD, Jackevicius CA, Marzec LN, et al. Adherence and outcomes to direct oral anticoagulants among patients with atrial fibrillation: findings from the veterans health administration. BMC Cardiovasc Disord 2017 Sep 2;17(1):236 [FREE Full text] [doi: 10.1186/s12872-017-0671-6] [Medline: 28865440]
11. Mekaj YH, Mekaj AY, Duci SB, Miftari EI. New oral anticoagulants: their advantages and disadvantages compared with vitamin K antagonists in the prevention and treatment of patients with thromboembolic events. Ther Clin Risk Manag 2015;11:967-977 [FREE Full text] [doi: 10.2147/TCRM.S84210] [Medline: 26150723]
12. Mayring P. Qualitative Inhaltsanalyse. In: Mey G, Mruck K, editors. Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: Vs Verlag; 2010:601-613.
13. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Fourth Edition. Thousand Oaks: Sage Publications Ltd; 2015.
14. Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Research. Second Edition. Thousand Oaks: Sage Publications Ltd; 2014.
15. Glaser BG, Strauss AL. The Discovery Of Grounded Theory: Strategies For Qualitative Research. Chicago: Aldine Publishing Company; 1967.

16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
17. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-7.e1. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]
18. Manteuffel M, Williams S, Chen W, Verbrugge RR, Pittman DG, Steinkellner A. Influence of patient sex and gender on medication use, adherence, and prescribing alignment with guidelines. J Womens Health (Larchmt) 2014 Feb;23(2):112-119. [doi: 10.1089/jwh.2012.3972] [Medline: 24206025]
19. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
20. Uchino BN. Social support and health: a review of physiological processes potentially underlying links to disease outcomes. J Behav Med 2006 Aug;29(4):377-387. [doi: 10.1007/s10865-006-9056-5] [Medline: 16758315]
21. Yao X, Abraham NS, Alexander GC, Crown W, Montori VM, Sangaralingham LR, et al. Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation. J Am Heart Assoc 2016 Feb 23;5(2):pii: e003074 [FREE Full text] [doi: 10.1161/JAHA.115.003074] [Medline: 26908412]
22. Kronish IM, Ye S. Adherence to cardiovascular medications: lessons learned and future directions. Prog Cardiovasc Dis 2013;55(6):590-600 [FREE Full text] [doi: 10.1016/j.pcad.2013.02.001] [Medline: 23621969]
23. Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. Circulation 2008 Feb 26;117(8):1028-1036. [doi: 10.1161/CIRCULATIONAHA.107.706820] [Medline: 18299512]
24. Anderson M, Perrin A. Pew Research Center. 2017 May 17. Tech Adoption Climbs Among Older Adults URL: http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2019-11-06] [WebCite Cache ID 73vadXF7h]
25. Barton SM. RMIT Research Repository. 2010 Sep. Social and Cultural Factors that Influence the Uptake of E-learning: Case Studies in Malaysia, Indonesia, Turkey, Singapore and Australia URL: https://researchbank.rmit.edu.au/eserv/rmit:9729/Barton.pdf [accessed 2019-11-08]
26. Tse T, Tiong J, Kangaslahti V. The effect of cultural norms on the uptake of Information and Communication Technologies in Europe: a conceptual analysis. Int J Manag 2014;21(3):382-392 [FREE Full text]
27. Al-Jumeily D, Hussain A. The impact of cultural factors on technology acceptance: a technology acceptance model across Eastern and Western cultures. Int J Enhanced Res Educ Dev 2014;2(4):37-62 [FREE Full text]
28. Alagöz F, Ziefle M, Wilkowska W, Valdez AC. Openness to accept medical technology - a cultural view. In: Holzinger A, Simonic KM, editors. Information Quality in e-Health. Berlin, Heidelberg: Springer; 2011:151-170.
29. Alsswey A, Naufal I, Bervell B. Investigating the acceptance of mobile health application user interface cultural-based design to assist Arab elderly users. Int J Adv Comput Sci Appl 2018;9(8):144-152. [doi: 10.14569/IJACSA.2018.090819]
30. Prochaska J, DiClemente C. Transtheoretical therapy: toward a more integrative model of change. Psychother 1982;19(3):276-288 [FREE Full text] [doi: 10.1037/h0088437]
31. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
32. Aldeer M, Javanmard M, Martin R. A review of medication adherence monitoring technologies. Appl Syst Innov 2018;1(2):14. [doi: 10.3390/asi1020014]
33. Shah V, Dileep A, Dickens C, Groo V, Welland B, Field J, et al. Patient-centered tablet application for improving medication adherence after a drug-eluting stent. Front Public Health 2016;4:272 [FREE Full text] [doi: 10.3389/fpubh.2016.00272] [Medline: 28018897]
34. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e18 [FREE Full text] [doi: 10.2196/mhealth.7030] [Medline: 28223265]
35. Browne SH, Behzadi Y, Littlewort G. Let visuals tell the story: medication adherence in patients with type II diabetes captured by a novel ingestion sensor platform. JMIR Mhealth Uhealth 2015 Dec 31;3(4):e108 [FREE Full text] [doi: 10.2196/mhealth.4292] [Medline: 26721413]
36. Mauro JA, Mathews K, Sredzinski E. Avella Specialty Pharmacy. 2016. Effect of a 'smart pill-bottle' on medication adherence in patients who have multiple myeloma and are new to lenalidomide therapy: interim analysis URL: https://www.avella.com/hubfs/Clinical_Poster_Pill_Bottle_Adherence.pdf [accessed 2019-11-06] [WebCite Cache ID 73vcKVtDL]
37. Kidd CD. Catalia Health. 2008 Feb. Designing for Long-term Human-robot Interaction and Application to Weight Loss URL: http://www.cataliahealth.com/wp-content/uploads/2015/10/KiddPhDThesis.pdf [accessed 2018-11-14] [WebCite Cache ID 73vdRUXvA]
38. DeMeo M, Morena M. Medication Adherence Using a Smart Pill Bottle. In: Proceedings of the 2014 11th International Conference & Expo on Emerging Technologies for a Smarter World. 2014 Presented at: CEWIT'14; October 29-30, 2014; Melville, NY, USA. [doi: 10.1109/CEWIT.2014.7021149]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
NOAC: new oral anticoagulants
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.12.18; peer-reviewed by M Kelley, H Oh, S Browne; comments to author 28.02.19; revised version received 17.07.19; accepted 26.09.19; published 24.01.20.

Please cite as:
Herrmann M, Boehme P, Hansen A, Jansson K, Rebacz P, Ehlers JP, Mondritzki T, Truebel H
Digital Competencies and Attitudes Toward Digital Adherence Solutions Among Elderly Patients Treated With Novel Anticoagulants: Qualitative Study
J Med Internet Res 2020;22(1):e13077
URL: http://www.jmir.org/2020/1/e13077/
doi:10.2196/13077
PMID:32012049

Original Paper

A Web-Based, Computer-Tailored Intervention to Reduce Alcohol Consumption and Binge Drinking Among Spanish Adolescents: Cluster Randomized Controlled Trial

José Manuel Martinez-Montilla[1,2,3], MSc; Liesbeth Mercken[2,3], PhD; Hein de Vries[2,3], PhD; Math Candel[3,4], PhD; Joaquín Salvador Lima-Rodríguez[1], PhD; Marta Lima-Serrano[1], PhD

[1]Department of Nursing, School of Nursing, Physiotherapy and Podiatry, University of Seville, Seville, Spain

[2]Department of Health Promotion, Maastricht University, Maastricht, Netherlands

[3]Care and Public Health Research Institute, Maastricht, Netherlands

[4]Department of Methodology and Statistics, Maastricht University, Maastricht, Netherlands

Corresponding Author:
José Manuel Martinez-Montilla, MSc
Department of Nursing
School of Nursing, Physiotherapy and Podiatry
University of Seville
Street Avenzoar nº 6
Seville, 41009
Spain
Phone: 34 669955295
Email: josmarmon3@alum.us.es

Abstract

Background: Alcohol consumption, including binge drinking (BD) and heavy episodic drinking (HED), is one of the leading risk factors among Spanish adolescents leading to significant social, health, and economic consequences. Reduction of BD and HED in adolescents can be achieved using Web-based, computer-tailored (CT) interventions, providing highly personalized feedback that is adapted to a person's individual characteristics and needs. Randomized controlled trials assessing the effects of tailored BD reduction programs among Spanish adolescents are scarce.

Objective: The aim of this study was to test the effectiveness of the Web-based, CT intervention Alerta Alcohol, aimed at the prevention of BD in Spanish adolescents. As a secondary outcome, effects on HED, weekly consumption, and any consumption were also assessed. The adherence and process evaluation were assessed.

Methods: A cluster randomized controlled trial conducted among 15 Spanish schools was developed. Each school was randomized into either an experimental condition (EC) (N=742) or a control condition (CC) (N=505). Finally, 351 participants for the EC and 261 for the CC were included in the analysis (N=612). Baseline assessment took place in January and February 2017. Demographic variables and alcohol use were assessed at baseline. Follow-up assessment of alcohol use took place 4 months later in May and June 2017. Participants were compared according to their randomization group (EC versus CC). After the baseline assessment, participants in the EC started the intervention, which consisted of short stories about BD, in which CT feedback was based on the I-Change Model for behavior change. Participants in the CC group only received the baseline questionnaire. Effects of the intervention were assessed using a three-level mixed logistic regression analysis for BD, HED, and any consumption, and a three-level mixed linear regression analysis for weekly consumption.

Results: In total, 1247 adolescents participated in the baseline assessment and 612 participated in the follow-up assessment; the attrition rate was 50.92%. The intervention was effective in reducing HED among adolescents; the odds of HED in the CC was nine times that in the experimental condition (P=.04). No effects were found for BD, weekly consumption, and any consumption. Process evaluations revealed that the adolescents were satisfied with the program (68.8%), would use the program again (52.9%), and would recommend it to someone else (62.8%). Females and non-binge drinkers showed better responses in the process evaluation.

Conclusions: Our intervention was effective regarding HED but not regarding BD, weekly consumption, and any consumption. It may be that limiting alcohol consumption to prevent HED was easier in the Spanish context than it was to carry out further

steps, such as reducing other patterns of alcohol consumption. Hence, additional actions are needed to accomplish these latter goals, including community approaches and policy actions aimed at denormalizing alcohol consumption among Spanish adolescents.

Trial Registration: ClinicalTrials.gov NCT03288896; https://clinicaltrials.gov/ct2/show/NCT03288896

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5346-4

(J Med Internet Res 2020;22(1):e15438) doi:10.2196/15438

KEYWORDS

adolescents; alcohol consumption; binge drinking; cluster randomized controlled trial; computer tailoring

Introduction

Alcohol consumption is one of the leading risk factors for mortality and disease worldwide [1,2], with significant social, health, and economic consequences [3], and is the leading risk factor globally for both death and disability-adjusted life years [4]. In addition, the most common risky drinking behavior among adolescents is binge drinking (BD), which consists of alcohol consumption that results in a blood alcohol concentration of 0.08 g/dL or more, within a period 2 hours [5,6]. In men, blood alcohol concentrations of more than 0.08 g/dL typically occur after consuming five or more standard drink units (SDUs), or standard glasses of alcohol, in about 2 hours; in women, this occurs after consuming four or more SDUs in about 2 hours [5,7]. In other words, BD occurs with the consumption of four or more and five or more SDUs of alcohol by women and men, respectively, in a short space of time or during a single occasion [5-8]. The lack of consensus as to what is considered an SDU could be the result of cross-country variability of criteria regarding the amount of alcohol consumption per episode [6,9,10]. However, the above definition, which is used in our study, is consistent with the Spanish epidemiological data from the survey Encuesta Sobre Uso de Drogas en Estudiantes de Enseñanzas Secundarias (ESTUDES) [8]. The fact is that BD is associated with detrimental long- and short-term consequences, since it affects neurocognitive development and leads to physical, psychological, and social alterations. In addition, BD has been associated with traffic accidents, violence, homicide, suicide, early sexual contact, school failure, mental illnesses, and delinquency, among other issues [3,11-15].

Moreover, heavy episodic drinking (HED) has been defined as the consumption of 10 or more glasses of alcohol on at least one occasion in the previous week [16,17]. This pattern of alcohol consumption is believed to be a serious problem in Western society, with major psychological, social, and economic consequences [17,18]. Similarly, HED has been linked to several problematic behaviors, such as an increased risk of unplanned sexual activity; increased risk of injury [19]; being more likely to engage in delinquent acts, including fighting, truancy from school, stealing, or driving while intoxicated[17]; as well as later alcohol abuse and dependence or illegal drug use [20].

The 2018 national Spanish survey ESTUDES showed that 75.6% of adolescents between 14 and 18 years of age drank alcohol in the last 12 months [8], a slightly lower number compared to that reported in the 2015 European School Survey Project on Alcohol and Other Drugs (ESPAD) [21]. Furthermore, in Europe, 35% of 15-16-year-old adolescents reported BD [21], while in Spain, the percentage stood at 32.2% of students between 14 and 18 years of age [8]. The national Spanish survey showed that the prevalence of BD at the age of 16 was 37.0%, which was twice that of 14-year-olds (13.2%), and BD reached 47.5% at the age of 17 [8]. Not one of the previous reports shows data on HED; however, Best et al [17] found a prevalence of 32% of HED among 14-16-year-old adolescents in the United Kingdom. These figures highlight the importance of preventing these different patterns of alcohol consumption in adolescents.

In the prevention area, computer- and Internet-based interventions are increasingly used as platforms for health promotion, including interventions aimed at reducing alcohol consumption [16,22-26]. Thus, the reduction of alcohol use and BD in adolescents could be achieved with the help of Web-based, computer-tailored (CT) interventions [16,27]; these could provide highly personalized feedback to individuals whose behaviors and opinions would be previously assessed on the basis of their answers to questionnaires, using data-driven decision rules that produce personalized feedback automatically from a database [28]. Some studies have shown that tailored advice helps to effectively change health behaviors and their determinants [16,23,24,27,29-31], even showing it to be cost-effective [32], although their effect sizes were generally small to medium [16,33]. However, Web-based, CT interventions usually have low adherence and high attrition rates [16,34,35], including over 50% [16,27,34,36], causing significant negative consequences, such as a reduced ability to reveal intervention effects [37].

The *Alerta Alcohol* program is the first Spanish program to be implemented that consists of a dynamic, Web-based, CT intervention in a school environment aimed at the prevention of alcohol consumption and BD in Spanish adolescents. This is a cultural adaptation of the Dutch program carried out by Jander et al [16,38]; however, in the Spanish context, no similar study has targeted these issues to date, and using CT technology at the high school level is still very rare [39]. Moreover, Alerta Alcohol tries to improve program adherence and minimize dropout in an attempt to overcome the limitations of previous studies. This program uses different strategies to accomplish this, such as developing a dynamic intervention with stories adapted to gender and age [23,39,40], based on the feedback from focus groups with adolescents (paper not yet published), or by carrying out the majority of the study at schools as part of the health promotion curriculum.

The aim of this study was to test the effectiveness of the Web-based, CT intervention Alerta Alcohol. We assessed the

effects of the intervention on *BD*, *HED*, *weekly consumption*, and *any consumption*.

Methods

Ethics Committee Approval

The intervention was carried out according to bioethical guidelines; the students needed to answer the questionnaires themselves and confidentiality was guaranteed. Active informed consent was used. The project was approved the Bioethical Committee of Andalusia, Spain, and was registered on August 4, 2015 (registration number: PI-0031-2014). In addition, this trial was retrospectively registered at ClinicalTrials.gov on September 19, 2017 (NCT03288896). The intervention was not registered prospectively because our organization did not require it. However, the intervention was not modified with respect to the study protocol.

Study Design

We conducted a cluster randomized controlled trial, with one experimental condition (EC) and one waiting-list control condition (CC); high schools were randomized into these two groups. Participants completed a baseline (ie, pretest) evaluation and a final (ie, posttest) evaluation performed 4 months after the intervention.

Participants were compared based on their randomization groups: EC versus CC. The EC group received the online intervention that contained CT feedback. The CC group only filled in the online baseline questionnaire. Both groups were given an online follow-up assessment after 4 months; they completed the same questionnaire that was used in the baseline assessment. The study took place in Spain between January and June 2017. The Consolidated Standards of Reporting Trials (CONSORT) guidelines were followed [41] (see Multimedia Appendix 1).

Participants and Procedure

Participants were randomly selected from the group of students belonging to the public school system; this group included students in their fourth year of compulsory secondary education (CSE), those in the first year of their baccalaureate program, and those in the first year of continuing education or vocational training (VT), which is equivalent to 10th, 11th, and 12th grades in the United States, respectively. The randomization process was undertaken by two researchers from the team (JMMM and MLS) using a computer software randomization device to avoid contamination. First, we randomly selected at least two schools from each of the eight provinces in Andalusia in Southern Spain. If schools agreed to participate, the inclusion criteria were checked. If they did not agree to participate, we randomized other schools in the same province until at least two schools in each province were included; in total, we contacted 37 Andalusian schools.

The schools were informed of the objective of the study and the sessions of the intervention. Participation by each school was confirmed by email, telephone, or, when necessary, by a visit. A formal letter and an information folder were sent to teachers and coordinators at each school, where they were provided with contact details and the study website address [42], as well as a manual with frequently asked questions that may occur during the program.

Finally, after 16 high schools—two from each province in Andalusia—accepted, they were randomly assigned to either the EC group or the CC group, taking care that the intervention groups (ie, EC groups) were matched with a province. Within each school, all classes that met the inclusion criteria were invited to participate in the study.

The CC schools were on a waiting list and received the intervention voluntarily once the study was completed. The selected schools were not blinded to their groups, since the EC group needed to schedule a total of four sessions during school hours. The adolescents were recruited from schools through their teachers and counselors. Adolescent participants and their parents had to sign and return the informed consent form to agree to take part in this scientific study. When starting the intervention, participants were asked to visit the study website and create an account. Within their account, they selected their school and were assigned to one of the conditions: CC or EC. Before starting with the baseline questionnaire, students gave informed consent by checking the acceptance box on the first page of the website. If he or she did not wish to participate, or refused to provide informed consent, he or she could select the option *I do not wish to participate in this study*. In this case, he or she was thanked and could leave the website. Participants could, however, also access the website at another time if they wanted to (see Figure 1). All students enrolled gave consent through active informed consent for the use of their data for scientific research and publication. Those who were underage were asked that their parents complete an informed consent form.

Figure 1. Screenshot of the informed consent page from the Alerta Alcohol website.

Inclusion and Exclusion Criteria

The target group for the Alerta Alcohol program consisted of adolescents aged 15-19 years old that were enrolled in fourth-year CSE, first-year baccalaureate programs, or first-year VT. All students enrolled had Internet access at schools and in their homes. Those with language difficulties or those who had previously participated in prevention programs of BD were excluded. To check the inclusion criteria, a researcher was present at the pretest.

In addition, the inclusion criteria for schools were as follows: (1) public secondary schools from Andalusia, (2) schools belonging to provincial capitals, and (3) schools with access to the Internet and an equipped information and communication technology room available for students.

Intervention

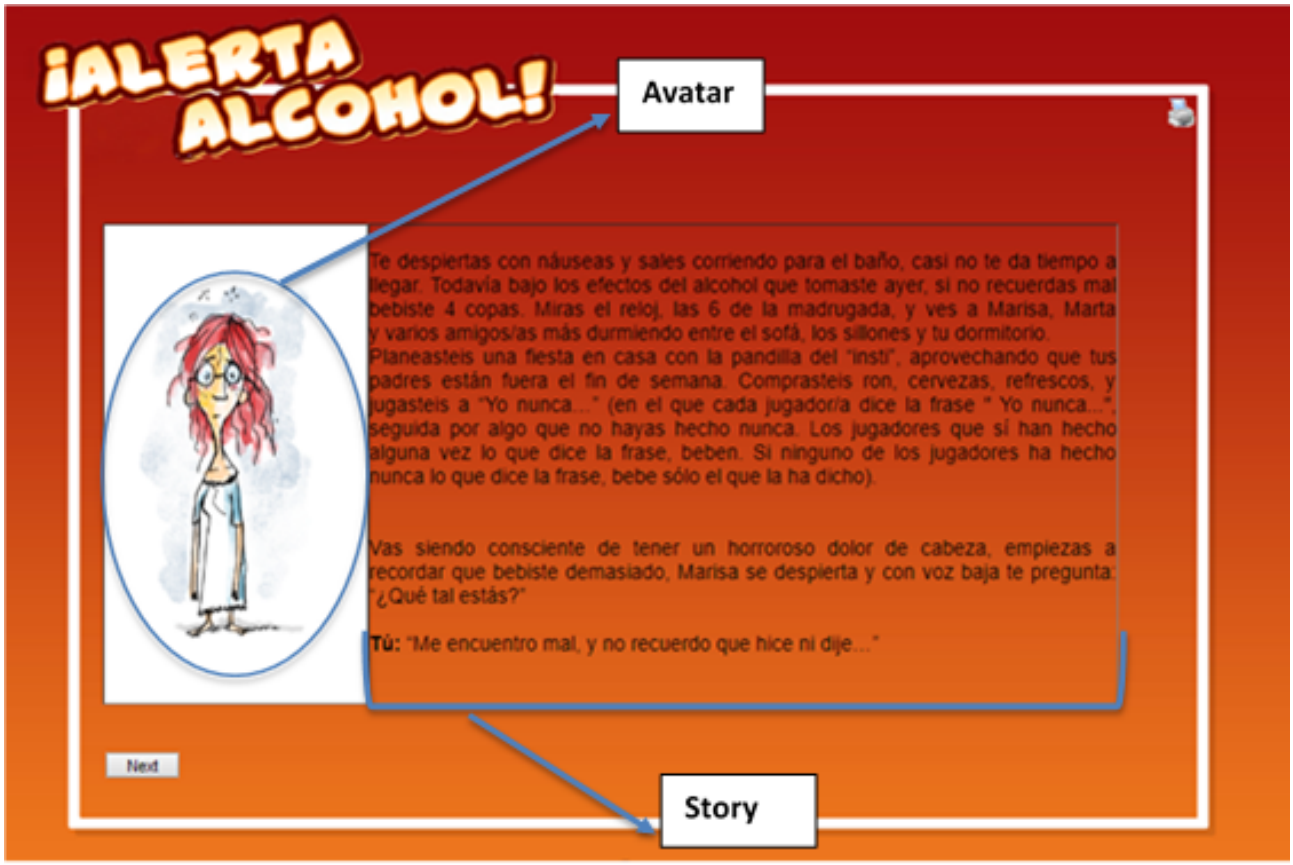
Alerta Alcohol consists of short stories in which the main character binge drank the night before and his or her friends talk with him or her about what happened the night before. The drinking event took place in three scenarios: at home, at a celebration, and in a public place. The stories were designed based on the results of a focus group study (paper not yet published) and were adapted to the gender of the participant. Participants could choose an avatar and the names of the characters in the stories (see Figures 2 and 3).

First, the stories were presented and questions and tailored messages were shown, which were designed to reduce alcohol consumption and BD. Concise, direct, and personalized relevant messages were delivered to promote participation in the intervention [16,38,40]. The messages were customized with the names of the participants; elements such as repetition of the answers were used to show respect and empathy, counter persuasion, introduce social modelling and new beliefs, and reinforce positive behaviors and motivational feedback [28,43]. These messages were based on the I-Change Model, in which the central concepts are attitude, social influences, self-efficacy, and action planning [39,44]. In each scenario, self-efficacy is reinforced and action plans are offered to the adolescent in order to reject alcohol and BD in the specific scenario. In addition, we developed questions and tailored messages aiming to increase self-esteem and awareness of factors such as the acknowledgement of risk perception of alcohol consumption and BD.

Based on a previous focus group study, we concluded that it was necessary to favor premotivational and motivational components associated with BD; modification of beliefs; and risk perception and expectations, or those associated with the effects of alcohol experimentation; as well as to promote healthy self-esteem and self-efficacy. Gender differences were also taken into account (paper not yet published). In addition, a Delphi group study for the cultural adaptation study and the pilot study was developed, which allowed us to shorten and rewrite the feedback messages to make them more appealing to our target group (paper not yet published).

Figure 2. Screenshot of the page for choosing an avatar from the Alerta Alcohol website.

Figure 3. Screenshot of example stories for a girl from the Alerta Alcohol website.



The intervention was carried out over six sessions. All students had access to a computer in their school's computer room. After registering themselves online, all students had to fill out a baseline questionnaire (ie, pretest), which was supervised by a

researcher. The following week, each student from the EC group continued with the intervention by logging in to the website; if needed, they were assisted by their teacher. The EC group members attended a second and third session held at school. There was a 1-2-week period between sessions; when possible, the schedule, which was planned with the researcher, was developed within the school curriculum (ie, class time). Students had two booster sessions at their homes 1 week after the third session. The fourth session was called *The Challenge*, where adolescents could accept the challenge of not drinking or at least not binge drink at an upcoming drinking event; the program reminded students again of the advice and action plans for this type of event. In the fifth session, 2 days after the drinking event, the program evaluated the challenge to determine whether or not the participants drank or at least whether they engaged in BD during this drinking event. At the end of the study, all students had to complete the sixth session at school (ie, the follow-up questionnaire), which was carried out throughout May and June 2017 (see Table 1). Each session took approximately 1 hour. The CC group only has two sessions: baseline data collection in January and February 2017 and the follow-up questionnaire (ie, posttest) 4 months later. A research technician was present for the pretest and posttest questionnaires at the schools to collaborate with the teachers and to optimize follow-up rates. Moreover, the researcher monitored the intervention by phone call or, if necessary, by visiting the schools in the EC group. Participation reminders were also sent via email when participants had not finished the intervention procedures, so they could complete them outside of school. Because of the nature of study, data were not anonymous. However, confidentiality was ensured through proper data management and security, according to the European General Data Protection Regulation (GDPR). This information was enclosed with the consent form. A detailed description of the development and content of the intervention is available in the protocol of this study [39].

Table 1. Alerta Alcohol program structure.

Location and session	Measures
At school	
1	Initial questionnaire
2	Scenario 1: at home • Knowledge and risks • Attitude: pros and cons • Self-efficacy and action plans
3	Scenario 2: celebrations • Self-esteem • Social modelling • Self-efficacy and action plans Scenario 3: public places • Social norms • Social pressures • Self-efficacy and action plans
At student's home	
4	Booster session: The Challenge
5	Evaluation of The Challenge
At school	
6	Final questionnaire

Measures

Overview

Multiple guidelines for how drinking should be measured in surveys have been proposed; however, whether they are consistent in their recommendations has not been considered to date [45]. In this study, a Spanish validated version of the self-administered online questionnaire was used [46], which was adapted from a previous study carried out on Dutch adolescents [38]. A better description of the concepts and variables can be found in the study protocol [39] and in Multimedia Appendix 2.

Demographics

Social demographic variables were assessed at baseline, which consisted of gender (1=male and 2=female), age (in years), educational level (1=CSE, 2=baccalaureate, and 3=VT), religion (1=Catholic, 2=Protestant/Evangelical, 3=Muslim/Islamic, 4=other religion, and 5=no religion), and ethnicity (1=Spanish and 2=other).

In addition, we used the Family Affluence Scale (FAS) to measure social status. The FAS consists of four different questions: "Does your family have a car or a van?" "Do you have your own room at home?" "During the last 12 months, how many times have you gone on holiday with your family?" and "How many computers does your family have?" [47]. The

FAS was transformed into three categories, where students could belong to the low level (0-2 points), middle level (3-5 points), or high level (6-9 points).

The Family Apgar Test was used to measure self-perception on familiar functional status (alpha=.778). It consists of five questions answered on a 3-point Likert scale (0=almost never, 1=sometimes, and 2=almost always); the resulting score was dichotomized into dysfunctional family status (score 0-6) and functional family status (score 7-10) [48].

Binge Drinking, Heavy Episodic Drinking, Weekly Consumption, and Any Consumption

BD was assessed using an open-ended question on how many BD occasions they participated in during the previous 30 days (eg, for girls, "How often did you drink 4 or more standard glasses of alcohol on one occasion in the previous 30 days?"; for boys, the number of drinks was 5 or more). A figure showing different standard drinks was shown to make the concept more comprehensible [16]. This variable was then dichotomized (0=reported no BD and 1=reported BD).

For HED, participants were dichotomized into two groups: those who consumed 10 or more glasses of alcohol on at least one occasion in the previous week (1=HED) and those who did not (0=no HED) [16,17].

For weekly consumption, we assessed how many glasses of alcohol students drank each day during the last week. Based on this information, we calculated the total number of glasses consumed in the past week [16].

Any consumption was calculated using the question "On which days of the past week did you drink alcohol?" Possible answers were as follows: "Monday to Sunday"; "I haven't drunk in the past week"; and "I have never drunk alcohol." This variable was dichotomized into two groups (0=no and 1=yes).

Process Evaluation and Adherence

To assess adherence, the number of intervention sessions attended by the participants at schools was registered. Furthermore, after completing each session, we asked respondents whether they have been reading the advice, whether the intervention was useful, whether the content was credible and appropriate, and whether the advice was interesting, understandable, too long or short, and personally relevant; responses were based on a 5-point Likert scale (eg, 1=totally disagree and 5=totally agree). We also assessed whether the advice increased their knowledge, changed their risk perception, changed their attitude, or improved their skills to prevent BD. For the analysis of these questions, answers were converted into three categories: 1=totally disagree/partially disagree, 2=neither agree nor disagree, and 3=totally agree/partially agree.

Finally, we also assessed general satisfaction with the program using a 5-point Likert scale. This question was converted into three categories: 1=very dissatisfied/dissatisfied, 2=neither satisfied nor dissatisfied, and 3=very satisfied/satisfied. Using a 5-point Likert scale, we also assessed whether, given the opportunity, they would use the program again and whether they would recommend the program to others (eg, 1=totally disagree and 5=totally agree) [49].

Power Analyses

The primary outcome was the difference in binge-drinking occasions in the previous 30 days in the EC group compared with the CC group. According to the 2016 national Spanish survey ESTUDES [50], the prevalence of adolescent BD within a previous 30-day timeframe was 32.2%. It is estimated that the intervention reduces consumption to 22%. Requiring a statistical power of .80 for a two-sided test with a type I error rate of alpha=.05, 309 subjects were required for the CC group and 309 were required for the EC group (ie, 618 participants in total); G*Power, version 3.1.9.2 (Heinrich-Heine-Universität Düsseldorf), was used for the statistical power analyses [51]. Following the study by Jander et al [38], a dropout rate of about 50% was anticipated; therefore, 1236 total subjects needed to be recruited.

Statistical Analyses

Descriptive analyses were performed to describe the characteristics of the participants at baseline. Differences between the conditions in the baseline sample, as well as between consumers and nonconsumers—BD, HED, weekly consumption, and any consumption—were assessed via a *t* test for continuous variables and a chi-square test for categorical variables. Also, when the dependent variable was not normally distributed, the Mann-Whitney *U* test was used.

Since pupils were nested within classes in the study, and classes were nested within schools, in order to examine predictors of dropout versus nondropout, a three-level mixed logistic regression analysis was conducted. To test the effectiveness of the program, we also performed three-level mixed logistic regression analyses for the outcomes *BD*, *any consumption*, and *HED*, and a three-level mixed linear regression for the outcome *weekly consumption*. When variances of the random intercept at the school level and class level turned out to be zero, a standard logistic or linear regression was carried out. This turned out to be the case for the binary outcomes *any consumption*, *BD*, and *HED*.

To evaluate the effect, the intention-to-treat principle was followed. The independent variable of interest was included in the EC versus CC, and covariates were the outcome at baseline as well as several sociodemographic variables: gender, age, educational level, religion, ethnicity, Apgar score, and affluence level. Also, the interaction effects between intervention condition and all sociodemographic variables were entered as covariates into the analyses. To build the refined model, we first examined whether the covariance model of random effects could be simplified. In the second phase, the variables with the least statistical significance, provided their significance was above .10 (for interaction terms) or .05 (for main effect terms), were eliminated one by one from the model. However, the variable of central interest—EC versus CC—always remained in the model. To quantify the predictive power of the logistic regression models, Nagelkerke's R^2 was reported. Due to the high attrition rate (>50%) in the majority of variables in the follow-up questionnaire, thus hampering the validity of multiple imputation techniques, we decided not to use multiple imputation and we performed the analysis with pairwise deletion [52,53].

To study predictors of adherence, we also analyzed the associations between potential participant characteristics (ie, gender, age, educational level, religion, ethnicity, Apgar score, family affluence, and alcohol use at baseline) on the one hand and participation in the intervention (ie, adherent or not) on the other.

Finally, for the process evaluation a descriptive analysis was performed using chi-square tests to examine differences between males and females and between binge drinkers and non-binge drinkers. We used SPSS Statistics for Windows, version 21.0 (IBM Corp), for these analyses. The level of significance used for the main effects was alpha=.05 and for the interaction effects was alpha=.10. Also, effect sizes (odds ratios [ORs]) and 95% CIs were calculated.

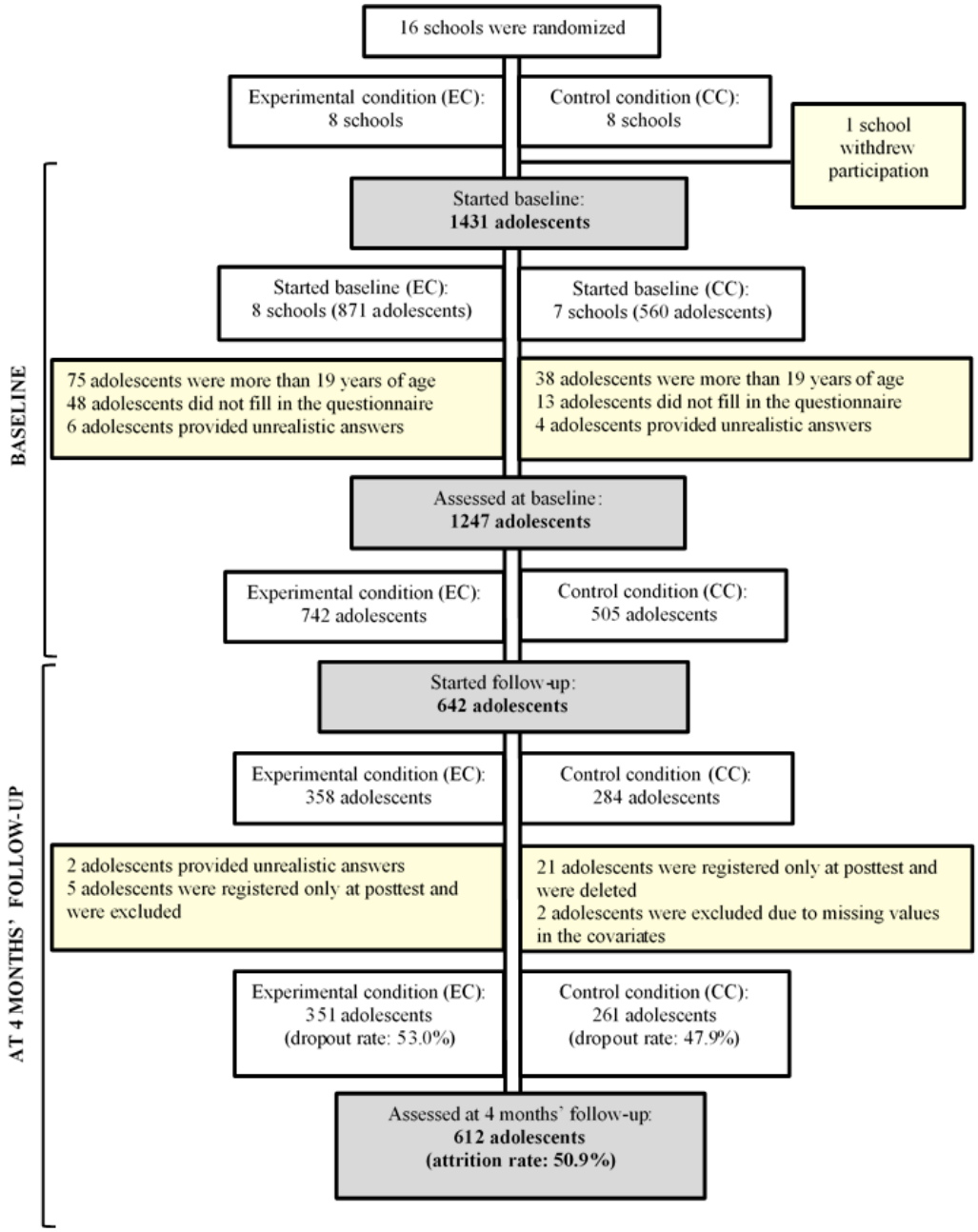
Results

Participation and Attrition Rate

After contacting several schools, 16 accepted our invitation. Reasons for nonacceptance were as follows: the program was too time-consuming, the schools were involved in another drug-related program, or there were logistical problems, such as not having a computer room. Other schools did not respond to our phone calls and emails.

In total, 16 schools were randomly assigned to either the EC or CC. However, one of the schools from the CC did not begin the baseline assessment and withdrew their participation due to logistical problems. In total, 1431 adolescents from 15 schools—8 EC and 7 CC—were requested to fill in the baseline questionnaire. Subsequently, 75 adolescents from the EC group and 38 from the CC group were removed because they were over 19 years of age, 48 students were removed from the EC group and 13 from the CC group because they did not fill in the questionnaire, and an additional 6 students from EC group and 4 from the CC group were removed because they provided unrealistic answers. In total, 1247 interviews were used at the baseline assessment—742 from the EC and 505 from the CC. Figure 4 shows the flowchart of participants at the participating schools.

Figure 4. Flowchart of participant recruitment and dropouts in the Alerta Alcohol program.



All 15 schools participated in the completion of the 4-month follow-up questionnaire, with only 612 participants: 351 from the EC group and 261 from the CC group. There was a clear decrease in participation between the baseline assessment (N=742) and the follow-up assessment (N=351) in the EC group, resulting in an attrition rate of 50.92% (612/1247). Most schools had problems with their Wi-Fi or reported trouble finding a suitable date, due to final exams. Also, groups participating in VT were beginning their placements in companies around this time, so this was also a cause of dropout.

The dropout analyses showed that adolescents returning the follow-up questionnaires were significantly younger (P<.001) and had lower educational levels (P<.001). The adolescents that returned the questionnaires were less likely to be binge drinkers (P=.02) and had a lower weekly consumption (P=.01) (see Table 2). When a multilevel analysis was carried out, the older participants (OR 1.319, 95% CI 1.101-1.581, P=.003) and those who reported a higher weekly consumption (OR 1.045, 95% CI 1.004-1.087, P=.03) were more likely to drop out of the intervention.

Table 2. Differences in baseline characteristics between participants that returned and those that dropped out at the 4-month follow-up.

Variable (number of missing values)	Total (N=1247)	Returned (n=612)	Dropped out (n=635)	Test statistic	*P* value
Age (15-19-year-olds) (0), mean (SD)	16.32 (1.07)	16.11 (0.99)	16.53 (1.10)	U^a=236280.5; $t_{1237.926}$=–7.090	<.001; <.001
Gender (0), n (%)				**χ^2_1=0.5**	**.49**
Male	586 (46.99)	282 (46.1)	304 (47.9)		
Female	661 (53.01)	330 (53.9)	331 (52.1)		
Educational level (0), n (%)				**χ^2_2=58.8**	**<.001**
Compulsory secondary education	545 (43.70)	313 (51.1)	232 (36.5)		
Baccalaureate program	515 (41.30)	252 (41.2)	263 (41.4)		
Vocational training	187 (15.00)	47 (7.7)	140 (22.1)		
Religion (3), n (%)				**χ^2_4=2.9**	**.57**
Catholic	767 (61.65)	385 (62.9)	382 (60.4)		
Protestant or Evangelical	26 (2.09)	11 (1.8)	15 (2.4)		
Muslim or Islamic	18 (1.45)	9 (1.5)	9 (1.4)		
Other	28 (2.25)	10 (1.6)	18 (2.8)		
No religion	405 (32.56)	196 (32.1)	209 (32.9)		
Ethnicity (3), n (%)				**χ^2_1=1.2**	**.26**
Spanish	1178 (94.69)	583 (95.4)	595 (94.0)		
Non-Spanish	66 (5.31)	28 (4.6)	38 (6.0)		
Apgar test (55), n (%)				**χ^2_1=1.2**	**.28**
Functional family	886 (74.33)	446 (75.7)	440 (73.0)		
Dysfunctional family	306 (25.67)	143 (24.3)	163 (27.0)		
Family affluence (49), n (%)				**χ^2_2=0.9**	**.63**
Low	33 (2.76)	16 (2.7)	17 (2.8)		
Medium	387 (32.30)	183 (31.0)	204 (33.5)		
High	778 (64.94)	391 (66.3)	387 (63.7)		
Alcohol use (0)					
Binge drinking, n (%)	487 (39.05)	219 (35.8)	268 (42.2)	χ^2_1=5.6	.02
Heavy episodic drinking, n (%)	16 (1.28)	8 (1.3)	8 (1.3)	χ^2_1=0.005	.95
Weekly consumption, mean (SD)	1.45 (4.30)	1.22 (3.33)	1.68 (5.06)	U=206367; $t_{1098.665}$=–1.929	.01; .054
Any consumption, n (%)	832 (66.72)	404 (66.0)	428 (67.4)	χ^2_1=0.4	.55

[a]Mann-Whitney *U* test.

Sample Characteristics

Baseline characteristics are shown in Table 3. We found that the adolescents in the EC group were older than those in the CC group (*P*<.001). We also found that the educational level of adolescents was higher in the CC group compared to the EC group (*P*=.002).

Table 3. Baseline characteristics of participants by condition.

Variable (number of missing values)	Total (N=1247)	Experimental group (n=742)	Control group (n=505)	Test statistic	*P* value
Schools, n	15	8	7	N/A[a]	N/A
Classrooms per school, n	62	38	24	N/A	N/A
Students per classroom, mean (SD), range	N/A	19.53 (6.90), 25	21.04 (5.87), 26	N/A	N/A
Age (15-19-year-olds) (0), mean (SD)	16.32 (1.07)	16.40 (1.06)	16.21 (1.06)	U[b]=208553; t_{1245}=–3.150	<.001; .002
Gender (0), n (%)				**χ^2_1=0.2**	**.67**
Male	586 (46.99)	345 (46.5)	241 (47.7)		
Female	661 (53.01)	397 (53.5)	264 (52.3)		
Educational level (0), n (%)				**χ^2_2=12.1**	**.002**
Compulsory secondary education	545 (43.70)	321 (43.3)	224 (44.4)		
Baccalaureate program	515 (41.30)	289 (38.9)	226 (44.8)		
Vocational training	187 (15.00)	132 (17.8)	55 (10.9)		
Religion (3), n (%)				**χ^2_4=3.1**	**.55**
Catholic	767 (61.65)	462 (62.3)	305 (60.6)		
Protestant or Evangelical	26 (2.09)	19 (2.6)	7 (1.4)		
Muslim or Islamic	18 (1.45)	10 (1.4)	8 (1.6)		
Other	28 (2.25)	15 (2.0)	13 (2.6)		
No religion	405 (32.56)	235 (31.7)	170 (33.8)		
Ethnicity (3), n (%)				**χ^2_1=1.9**	**.17**
Spanish	1178 (94.69)	707 (95.4)	471 (93.6)		
Non-Spanish	66 (5.31)	34 (4.6)	32 (6.4)		
Apgar test (55), n (%)				**χ^2_1=2.2**	**.14**
Functional family	886 (74.33)	516 (72.8)	370 (76.6)		
Dysfunctional family	306 (25.67)	193 (27.2)	113 (23.4)		
Family affluence (49)				**χ^2_2=3.5**	**.17**
Low	33 (2.76)	17 (2.4)	16 (3.3)		
Medium	387 (32.30)	244 (34.2)	143 (29.5)		
High	778 (64.94)	452 (63.4)	326 (67.2)		
Alcohol use (0)					
Binge drinking, n (%)	487 (39.10)	298 (40.2)	189 (37.4)	**χ^2_1=0.9**	**.33**
Heavy episodic drinking, n (%)	16 (1.30)	8 (1.1)	8 (1.6)	χ^2_1=0.6	.44
Weekly consumption, mean (SD)	1.45 (4.30)	1.47 (3.78)	1.42 (4.96)	U=190393.5	.52
				t_{1245}=-0.2	.84
Any consumption, n (%)	832 (66.70)	486 (65.5)	346 (68.5)	χ^2_1=1.2	.27

[a]Not applicable.

[b]Mann-Whitney *U* test.

Binge Drinking

At baseline, 40.2% (298/742) of adolescents in the EC group and 37.4% (189/505) in the CC group reported BD in the previous 30 days. At the follow-up session, 32.2% (113/351) of adolescents in the EC group and 33.0% (86/261) in the CC group reported BD. There was a reduction in BD in both conditions, but this reduction was only significant in the EC group: BD was reduced by 8.0% (OR 0.716, 95% CI 0.547-0.937, *P*=.02) in the EC group versus a reduction of 4.4% (OR 0.821, 95% CI 0.600-1.126, *P*=.22) in the CC group. However, in the logistic regression analysis, the intervention did not show a significant effect; although, in the CC group, the odds of BD were 1.1 times the odds of BD in the EC group. In addition, the analysis revealed that family affluence was marginally significant (*P*=.08); people who had medium family affluence (OR 3.365, 95% CI 1.058-10.704, *P*=.04) had a higher risk of BD than people who had high family affluence (see Tables 4 and 5).

Table 4. Effects of the Alerta Alcohol program on binge drinking, heavy episodic drinking, weekly consumption, and any consumption in the complete model.

Variable	Binge drinking[a] (N=586)		Heavy episodic drinking[a] (N=612)		Weekly consumption[b] (N=589)			Any consumption[a] (N=589)	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	B (SE)	*P* value	95% CI	OR (95% CI)	*P* value
Condition (control vs experimental)	1.106 (0.730-1.674)	.63	9.129 (1.107-75.259)	.04	–0.244 (0.385)	.53	–1.000-0.511	0.866 (0.506-1.484)	.60
Gender (male vs female)	N/A[d]	N/A	3.394 (0.669-17.223)	.14	N/A	N/A	N/A	N/A	N/A
Family affluence									
Low versus high	3.365 (1.058-10.704)	.04	N/A	N/A	–1.503 (0.921)	.10	–3.313-0.306	3.401 (0.710-16.129)	.13
Medium versus high	1.309 (0.841-2.036)	.23	N/A	N/A	0.410 (0.325)	.21	–0.228-1.048	1.801 (0.987-3.289)	.06
Consumption behavior pretest[e]	11.986 (7.951-18.069)	<.001	8.360 (0.820-85.239)	.07	0.502 (0.044)	<.001	0.416-0.589	59.518 (34.789-101.823)	<.001

[a]Logistic regression.

[b]Linear mixed regression.

[c]OR: odds ratio.

[d]Not applicable.

[e]Consumption behavior—binge drinking, heavy episodic drinking, weekly consumption, and any consumption—at the pretest evaluation.

Table 5. Fixed effects of the Alerta Alcohol program on binge drinking, heavy episodic drinking, weekly consumption, and any consumption in the complete model.

Variable	Binge drinking[a] (N=586)		Heavy episodic drinking[a] (N=612)		Weekly consumption[b] (N=589)		Any consumption[a] (N=589)	
	Statistic	*P* value	Statistic	*P* value	Statistic	*P* value	Statistic	*P* value
Condition	W[c]=0.225	.64	W=4.222	.04	F[d]=0.403	.53	W=0.271	.60
Family affluence	W=5.108	.08	N/A[e]	N/A	F=2.405	.09	W=5.443	.07
	R^{2f}=.349	N/A	R^2=.150	N/A	N/A	N/A	R^2=.638	N/A
Intracluster correlation (ICC)								
School level	N/A	N/A	N/A	N/A	ICC=0	N/A	N/A	N/A
Class level	N/A	N/A	N/A	N/A	ICC=.046	N/A	N/A	N/A

[a]Logistic regression.

[b]Linear mixed regression.

[c]Wald test.

[d]F test.

[e]Not applicable.

[f]Nagelkerke's R^2.

Heavy Episodic Drinking

At baseline, 1.1% (8/742) of adolescents in the EC group and 1.6% (8/505) in the CC group reported heavy episodic drinking in the previous 30 days. At the follow-up stage, 0.3% (1/351) of adolescents in the EC group and 2.7% (7/261) in the CC group reported heavy episodic drinking in the previous 30 days. The logistic regression analysis showed that the odds of heavy episodic drinking in the CC group was nine times the odds of heavy episodic drinking in the EC group (OR 9.129, 95% CI 1.107-75.259, *P*=.04).

Weekly Consumption

At baseline, adolescents in the EC group and in the CC group drank a mean of 1.47 (SD 3.78) and 1.42 (SD 4.96) standard glasses of alcohol, respectively, in the previous week. At the follow-up stage, the mean was 1.64 (SD 3.66) and 1.39 (SD 4.12) standard glasses of alcohol in the EC group and in the CC group, respectively. The mixed linear regression analysis did not show any significant effects of the intervention.

Any Consumption

At baseline, 65.5% (486/742) of adolescents in the EC group and 68.5% (346/505) in the CC group reported that they had never drunk alcohol. At the follow-up stage, 68.9% (242/351) of adolescents in the EC group and 65.5% (171/261) in the CC group reported that they had never drunk alcohol. In the logistic regression analysis, the intervention had no effect. Again, family affluence was marginally significant (*P*=.07), where participants with low family affluence (OR 1.801, 95% CI 3.289-0.987, *P*=.06) had a higher probability of drinking alcohol than those with high family affluence.

Adherence

After the baseline (ie, first) session, of the 742 adolescents who were randomized into the EC, only 461 (62.1%) started the second session at school and only 350 (47.2%) returned for the third session at school. Only 23 adolescents out of 742 (3.1%) returned for the fourth session at home and 8 (1.1%) returned for the fifth session at home.

We assessed the predictors of adherence, which were as follows: educational level (*P*=.009) (with students enrolled in VT showing less adherence than baccalaureate program students: beta=–0.882, *P*=.002), ethnicity (Spanish versus other: beta=1.142, *P*=.04), and not engaging in BD in the previous 30 days at baseline (beta=–0.546, *P*=.03) (see Table 6).

Table 6. Predictors of adherence where both sessions were completed.

Variable[a]	Beta	SE	Exp(beta)	*P* value
Gender (female)	0.319	0.229	1.376	.16
Age	–0.140	0.138	0.870	.31
Educational level (baccalaureate)				**.009**
Vocational training	–0.882	0.287	0.414	.002
Compulsory secondary education	–0.201	0.340	0.818	.55
Religion (no religion)[b]				**.30**
Others	0.005	0.256	1.005	.98
Catholic	0.950	0.627	2.586	.13
Ethnicity (others)	1.142	0.566	3.132	.04
Apgar test (family dysfunction)	0.162	0.287	1.176	**.57**
High family affluence				.19
Medium family affluence	1.313	1.091	3.717	.23
Low family affluence	0.385	0.257	1.470	.13
Binge drinking	–0.546	0.250	0.579	.03

[a]Reference category of categorical variables is indicated between brackets.

[b]Religion was entered as three categories: no religion, others, and Catholic.

Process Evaluation

Of the EC group members, 295 participants returned the questionnaire. In total, 50.8% (150/295) of students reported that the sessions were too long. Even though 76.1% (223/293) of students said the advice content was credible, 72.0% (211/293) stated that it was understandable, 63.1% (185/293) stated that it was useful, and 60.0% (177/295) stated that the advice was interesting. Furthermore, 68.8% (203/295) of students were satisfied with the program, 52.9% (155/293) would use the program again, and 62.8% (184/293) recommended the program to someone else (see Tables 7 and 8).

Table 7. Process evaluation of the Alerta Alcohol program: responses from participants in the intervention group by gender.

Variable (number of missing values)	Total (N=351), n (%)	Male (N=149), n (%)	Female (N=202), n (%)	χ^2_2	*P* value
Overall satisfaction (56)				**1.0**	**.61**
Very dissatisfied/dissatisfied	35 (11.9)	18 (14.0)	17 (10.2)		
Neither satisfied nor dissatisfied	57 (19.3)	25 (19.4)	32 (19.3)		
Very satisfied/satisfied	203 (68.8)	86 (66.7)	117 (70.5)		
Length of session: too long (56)				**2.4**	**.31**
Totally disagree/partially disagree	49 (16.6)	19 (14.7)	30 (18.1)		
Neither agree nor disagree	96 (32.5)	48 (37.2)	48 (28.9)		
Totally agree/partially agree	150 (50.8)	62 (48.1)	88 (53)		
Length of session: too short (56)				**1.0**	**.60**
Totally disagree/partially disagree	88 (29.8)	38 (29.5)	50 (30.1)		
Neither agree nor disagree	101 (34.2)	48 (37.2)	53 (31.9)		
Totally agree/partially agree	106 (35.9)	43 (33.3)	63 (38.0)		
Interest in advice (56)				**1.0**	**.60**
Totally disagree/partially disagree	52 (17.6)	26 (20.2)	26 (15.7)		
Neither agree nor disagree	66 (22.4)	28 (21.7)	38 (22.9)		
Totally agree/partially agree	177 (60.0)	75 (58.1)	102 (61.4)		
Content credibility (58)				**4.5**	**.10**
Totally disagree/partially disagree	24 (8.2)	15 (11.7)	9 (5.5)		
Neither agree nor disagree	46 (15.7)	22 (17.2)	24 (14.5)		
Totally agree/partially agree	223 (76.1)	91 (71.1)	132 (80.0)		
Advice was useful (58)				**6.9**	**.03**
Totally disagree/partially disagree	44 (15.0)	27 (21.1)	17 (10.3)		
Neither agree nor disagree	64 (21.8)	28 (21.9)	36 (21.8)		
Totally agree/partially agree	185 (63.1)	73 (57.0)	112 (67.9)		
Advice was understandable (58)				**>9.1**	**.01**
Totally disagree/partially disagree	25 (8.5)	18 (14.1)	7 (4.2)		
Neither agree nor disagree	57 (19.5)	25 (19.5)	32 (19.4)		
Totally agree/partially agree	211 (72.0)	85 (66.4)	126 (76.4)		
Advice was appropriate (58)				**4.3**	**.12**
Totally disagree/partially disagree	49 (16.7)	28 (21.9)	21 (12.7)		
Neither agree nor disagree	73 (24.9)	30 (23.4)	43 (26.1)		
Totally agree/partially agree	171 (58.4)	70 (54.7)	101 (61.2)		
Improved knowledge (58)				**9.5**	**.009**
Totally disagree/partially disagree	46 (15.7)	29 (22.7)	17 (10.3)		
Neither agree nor disagree	72 (24.6)	25 (19.5)	47 (28.5)		
Totally agree/partially agree	175 (59.7)	74 (57.8)	101 (61.2)		
Changed attitude (58)				**4.1**	**.13**
Totally disagree/partially disagree	54 (18.4)	30 (23.4)	24 (14.5)		
Neither agree nor disagree	90 (30.7)	39 (30.5)	51 (30.9)		
Totally agree/partially agree	149 (50.9)	59 (46.1)	90 (54.5)		
Changed risk perception (58)				**5.7**	**.06**
Totally disagree/partially disagree	49 (16.7)	29 (22.7)	20 (12.1)		

Variable (number of missing values)	Total (N=351), n (%)	Male (N=149), n (%)	Female (N=202), n (%)	χ^2_2	*P* value
Neither agree nor disagree	81 (27.6)	33 (25.8)	48 (29.1)		
Totally agree/partially agree	163 (55.6)	66 (51.6)	97 (58.8)		
Improved skills to reduce binge drinking (58)				**7.6**	**.02**
Totally disagree/partially disagree	45 (15.4)	28 (21.9)	17 (10.3)		
Neither agree nor disagree	86 (29.4)	33 (25.8)	53 (32.1)		
Totally agree/partially agree	162 (55.3)	67 (52.3)	95 (57.6)		
Would use the program again (58)				**5.2**	**.07**
Totally disagree/partially disagree	49 (16.7)	28 (21.9)	21 (12.7)		
Neither agree nor disagree	89 (30.4)	33 (25.8)	56 (33.9)		
Totally agree/partially agree	155 (52.9)	67 (52.3)	88 (53.3)		
Would recommend the program to someone else (58)				**6.1**	**.047**
Totally disagree/partially disagree	37 (12.6)	22 (17.2)	15 (9.1)		
Neither agree nor disagree	72 (24.6)	25 (19.5)	47 (28.5)		
Totally agree/partially agree	184 (62.8)	81 (63.3)	103 (62.4)		

Regarding gender, more females, compared to males, reported that the advice was useful (*d*=0.294, 95% CI 0.062-0.527, *P*=.03), understandable (*d*=0.316, 95% CI 0.083-0.548, *P*=.01), improved their knowledge (*d*=0.212, 95% CI 0.019-0.443, *P*=.009), changed their risk perception (*d*=0.236, 95% CI 0.005-0.468, *P*=.06), and improved their skills (*d*=0.229, 95% CI 0.003-0.460, *P*=.02). In general, more females, compared to males, more frequently reported that they would recommend the program to someone else (*d*=0.102, 95% CI 0.129-0.333, *P*=.047); this was a mixed result, since males also more often answered *totally agree/partially agree* for this question, as well as there being more males who answered *totally disagree/partially disagree*. Therefore, we could say that males are somewhat more extreme regarding their opinion on this question.

Regarding BD, more non-binge drinkers reported the advice to be credible (*d*=0.199, 95% CI 0.039-0.437, *P*=.03). On the other hand, more binge drinkers reported the advice to be useful (*d*=0.040, 95% CI 0.197-0.277, *P*=.02), to be appropriate for them (*d*=0.293, 95% CI 0.054-0.531, *P*=.01), and to improve their knowledge (*d*=0.068, 95% CI 0.170-0.305, *P*=.02), as well as that the advice changed their risk perception (*d*=0.078, 95% CI 0.159-0.315, *P*=.02).

Table 8. Process evaluation of the Alerta Alcohol program: responses from participants in the intervention group by binge-drinking status.

Variable (number of missing values)	No binge drinking (N=198), n (%)	Binge drinking (N=153), n (%)	χ^2_2	*P* value
Overall satisfaction (56)			**0.7**	**.69**
Very dissatisfied/dissatisfied	24 (12.9)	11 (10.1)		
Neither satisfied nor dissatisfied	34 (1.3)	23 (21.1)		
Very satisfied/satisfied	128 (68.8)	75 (68.8)		
Length of session: too long (56)			**5.9**	**.053**
Totally disagree/partially disagree	37 (19.9)	12 (11.0)		
Neither agree nor disagree	53 (28.5)	43 (39.4)		
Totally agree/partially agree	96 (54.6)	54 (49.5)		
Length of session: too short (56)			**6.0**	**.049**
Totally disagree/partially disagree	63 (33.9)	25 (22.9)		
Neither agree nor disagree	55 (29.6)	46 (42.2)		
Totally agree/partially agree	68 (36.6)	38 (34.9)		
Interest in advice (56)			**1.7**	**.43**
Totally disagree/partially disagree	36 (19.4)	16 (14.7)		
Neither agree nor disagree	38 (20.4)	28 (25.7)		
Totally agree/partially agree	112 (60.2)	65 (59.6)		
Content credibility (58)			**7.3**	**.03**
Totally disagree/partially disagree	15 (8.1)	9 (8.3)		
Neither agree nor disagree	21 (11.4)	25 (23.1)		
Totally agree/partially agree	149 (80.5)	74 (68.5)		
Advice was useful (58)			**7.6**	**.02**
Totally disagree/partially disagree	33 (17.8)	11 (10.2)		
Neither agree nor disagree	32 (17.3)	32 (29.6)		
Totally agree/partially agree	120 (64.9)	65 (60.2)		
Advice was understandable (58)			**2.4**	**.30**
Totally disagree/partially disagree	17 (9.2)	8 (7.4)		
Neither agree nor disagree	31 (16.8)	26 (24.1)		
Totally agree/partially agree	137 (74.1)	74 (68.5)		
Advice was appropriate (58)			**8.7**	**.01**
Totally disagree/partially disagree	40 (21.6)	9 (8.3)		
Neither agree nor disagree	43 (23.2)	30 (27.8)		
Totally agree/partially agree	102 (55.1)	69 (63.9)		
Improved knowledge (58)			**7.7**	**.02**
Totally disagree/partially disagree	35 (18.9)	11 (10.2)		
Neither agree nor disagree	37 (20.0)	35 (32.4)		
Totally agree/partially agree	113 (61.1)	62 (57.4)		
Changed attitude (58)			**2.3**	**.31**
Totally disagree/partially disagree	38 (20.5)	16 (14.8)		
Neither agree nor disagree	52 (28.1)	38 (35.2)		
Totally agree/partially agree	95 (51.4)	54 (50.0)		
Changed risk perception (58)			**7.6**	**.02**

Variable (number of missing values)	No binge drinking (N=198), n (%)	Binge drinking (N=153), n (%)	χ^2_2	*P* value
Totally disagree/partially disagree	34 (18.4)	15 (13.9)		
Neither agree nor disagree	41 (22.2)	40 (37.0)		
Totally agree/partially agree	110 (59.5)	53 (49.1)		
Improved skills to reduce binge drinking (58)			**3.8**	**.15**
Totally disagree/partially disagree	30 (16.2)	15 (13.9)		
Neither agree nor disagree	47 (25.4)	39 (36.1)		
Totally agree/partially agree	108 (58.4)	54 (50.0)		
Would use the program again (58)			**2.9**	**.23**
Totally disagree/partially disagree	34 (18.4)	15 (13.9)		
Neither agree nor disagree	50 (27.0)	39 (36.1)		
Totally agree/partially agree	101 (54.6)	54 (50.0)		
Would recommend the program to someone else (58)			**4.2**	**.12**
Totally disagree/partially disagree	27 (14.6)	10 (9.3)		
Neither agree nor disagree	39 (21.1)	33 (30.6)		
Totally agree/partially agree	119 (64.3)	65 (60.2)		

Discussion

Principal Findings

In this paper, the effectiveness of the first Web-based, CT intervention to reduce BD in Spanish adolescents was tested through a cluster randomized controlled trial. An overall effect of the intervention on BD was not found at 4 months, but there was an effect on reducing HED. In addition, no effects were found on weekly consumption or on any consumption.

In a similar previous study, Jander et al [16] found a significant link between condition and age, where their intervention turned out to be effective in reducing BD in adolescents between 15 and 16 years or age. In our study, although there was a trend in BD reduction in both conditions, effects of the intervention on BD behavior was not found. Nevertheless, the intervention was effective in reducing HED, in contrast to Jander et al [16], where this pattern of alcohol consumption showed a lower prevalence than in previous studies [16,54]. One explanation for our results may be that alcohol consumption is accepted in the Spanish context, including BD, whereas the more extreme type of drinking is less accepted and, thus, easier to change. As alcohol use is quite normative among Spanish adolescents [6,8,55], mostly in the context of parties or celebrations, it could be that the adolescents had no incentive to change their behaviors at these events [16,56]. In fact, in a previous qualitative study, many adolescents reported that they usually drink when they are at parties or during special occasions, such as fairs, Holy Week, or Christmas, showing a low-risk perception of BD (paper not yet published).

Moreover, in Spain, the alcohol drinking phenomenon called *Botellón* is one of the main sources of entertainment among young people, especially during nonschool periods, and it is normal for youths to binge drink at this kind of event [55,57]. Because of this, our study may have reduced the number of glasses of alcohol in the EC to be effective on HED, but it was not reduced enough to be effective on BD. Hence, future research needs to look at how to change norms regarding BD and how to prevent BD using more comprehensive campaigns and addressing cultural norms on BD. In addition, due to a lack of consensus as to what is considered an SDU [6,9,10], there may be varied results between countries, so these results may not be applicable to other populations.

Another possible explanation is that the follow-up in the EC, with respect to the CC, was closer to holiday periods, such as Holy Week or fairs, when adolescents usually go out to drink. Thus, this might be another reason related to the lack of effect on BD. Moreover, the intervention was only assessed in the fourth month; it would be advisable to carry out long-term assessments (ie, 12 or 24 months or an even longer time period) because several authors state that it requires a longer time period to see a real behavioral change [16,24].

Previous studies on CT interventions found high attrition rates [16,34,36], often caused by a lack of face-to-face contact and a high degree of anonymity [37,58]. A 50% attrition rate was taken into account in our power calculation, as with another study with a similar purpose to ours [39], and our study reported a lower attrition rate than that in other studies, such as Jander et al [16], Elfeddali et al [34], and Stanczyk et al [59], whose dropout rates were 68.9%, 62.9%, and 52.60%, respectively. Hence, strategies such as carrying out most of the sessions within the schools as part of the health promotion curriculum, collaborating with teachers at the pre- and posttest sessions by phone call, or, if necessary, assisting teachers at the schools, could be useful to minimize the dropout rate. That is why we believe that future studies should continue to carry out the intervention within the school curriculum to reduce dropout rates. Furthermore, older adolescents and those who were enrolled in higher educational-level programs dropped out more

often. Moreover, those who dropped out were more likely to engage in BD and reported a higher weekly consumption. These characteristics are consistent with other studies [16,27,34,60].

The recommendations by Jander et al [16] on how to approach the intervention were taken into account in the Alerta Alcohol program in trying to improve the adherence to the intervention, which seems to show slightly better results in our study. We focused on premotivational determinants, such as knowledge and risk perception [44,61], since, in a previous qualitative study, we found that the majority of adolescents were in the precontemplation phase—they had no intention of changing their health behavior (paper not yet published). According to the Transtheoretical Model, an increase of consciousness is an important first step leading toward behavior change [62]. We also developed a dynamic intervention, with different stories adapted to gender and age [23,39]. Moreover, we followed up on the intervention at the schools (eg, by phone call or, if necessary, in-person visit). It is possible that these actions could slightly improve the adherence to the Alerta Alcohol program.

Although two booster sessions were introduced at home, there was little participation; a possible solution to increase participation may be to incorporate these booster sessions within the school. Most of the participants stated that the reminder emails arrived in their spam folders, making the emails ineffective, which explains the low participation in these sessions. Some authors highlight the importance of the content and the schedule of the reminders [63,64], as well as sending notifications by other means, such as by text message and WhatsApp, among others, so these aspects could be improved in future studies.

To improve adherence in future interventions, it is necessary to be familiar with the predictors of adherence. In our study, adolescents belonging to VT and CSE groups showed less adherence. In this sense, it should be emphasized that the adolescents who belonged to the VT group could have had a higher dropout rate because they finished the academic year before the posttest, and at the time of the posttest they were out of school. This should be taken into account for upcoming interventions. Furthermore, we found that being Spanish also appears to be a predictor of adherence. Also, the analyses of adherence indicated that non-binge drinkers adhered better to the intervention, this last finding being typical in health promotion. This is because, as a rule, people who adhere better to a health program tend to have better lifestyle habits [65]. However, other studies, such as that conducted by Schneider et al [66], found that people with an unhealthy lifestyle usually visit their health intervention website more frequently; although, it is true that people with a healthier lifestyle were more likely to complete the health intervention. The problem is that health promotion programs should focus, particularly, on improving the health and lifestyle of people who lack a healthy lifestyle, even though the people who already lead a healthy lifestyle will strengthen their own. That is why we believe that more research is needed to know how to improve the adherence of people engaging in BD.

Finally, in the process evaluation, we found that adolescents in the EC group, in general, were very satisfied with the program. Regarding BD, binge drinkers were more satisfied with the program and showed a better response. Regarding gender, we can verify that females reported better a response to the program, which could be positive, since girls reported a higher rate of BD, and the intention of the health program is to focus on those who have an unhealthy lifestyle [66]. Moreover, it could be that boys are more reluctant to change their behavior or participate in programs to improve their health.

Strength and Limitations

The main strength of this study is that it was based on the I-Change Model for predicting healthy behavior acquisition and was preceded by quantitative [46] and qualitative research [39]. Furthermore, the Alerta Alcohol program was an adaptation of a previous intervention developed in the Netherlands based on extensive research that showed cost-effectiveness and effectiveness on BD among 15-16-year-olds [16,25,38]. Although several interventions exist regarding alcohol, few address BD, and even fewer interventions use CT technology in this target group at the school level. In addition, this study builds on earlier work done in the Netherlands, which was adapted to the Spanish context. As both replication and implementation are important for science, we are convinced that this intervention is highly relevant and innovative. Moreover, for the Spanish context, there is no similar study targeting alcohol consumption and BD prevention using a Web-based, CT intervention in adolescents. Moreover, our sample was randomized into EC and CC groups, and there were hardly any differences at baseline between these groups.

One of the limitations of this study was that only short-term outcomes of the intervention were assessed. It would be advisable to add more long-term assessments to evaluate the effects in the long run. In addition, the follow-up evaluation coincided with periods after holidays, which may have increased the probability of BD, as well as with periods in which adolescents belonging to the VT group were out of school, so this population was underrepresented. Further, many adolescents reported that some advice messages were long, therefore, they may have found them tiresome to read. In addition, it should be noted that Jander et al [16] developed their intervention in the context of a serious tailored game, which was not possible in our case, even though our program was a tailored dynamic intervention. Instead of a serious game, we used stories based on real facts identified in the previous focus group study.

Another limitation is that the sample size was determined for logistic regression, not taking clustering due to schools and classes into account. However, although we allowed for clustering in the analyses, for all binary outcomes the variances of the random effects turned out to be zero, which actually reduced the analysis to a logistic regression. Furthermore, it must be noted that we were rather close to complying with the required number of people—612 versus 618.

Moreover, compliance with the intervention protocol was relatively high for the initial classroom sessions, but this tailed off toward the end of the experimental period. However, this study improved the attrition rate through an adaptation of recommendations from the previous intervention developed in the Netherlands, resulting in a lower attrition rate—50.9% vs

68.9% [16]; high attrition rates are common in Web-based, CT interventions [67].

The large volume of missing data (>50%) was mainly due to missing values on the outcome variables, with less than 4% being due to missing values on covariates. In the case of missing values on outcomes, it is known that complete case analysis and multiple imputation yield the same results [52]. Since the fraction of missing cases due to covariates is very small, the added value of multiple imputation of these covariates is considered to be minor. It is also worth noting that, although multiple imputation might be of added value, there is also the risk of choosing an incorrect imputation model, which could lead to biased results [53].

Finally, one should be cautious when interpreting these results on the effectiveness of the program for the variable *HED*, since only 16 subjects were classified as heavy drinkers at the pretest stage.

Conclusions

We observed that the rate of BD decreased more in the EC group than in the CC group, and the overall effect on BD was not significant. However, we did find that there was an effect on HED. We believe that using the CT intervention at schools could be a good option for reducing some alcohol patterns such as HED among adolescents. However, we must be cautious when interpreting the results due to the low number of subjects in the HED group, which could affect the generalization of the study. Further research with long-term measurements is needed, as well as improvements in adherence to eHealth interventions, which will increase effectiveness and public health impact. In this sense, as one of the strategies for improving the program was to shorten the messages, perhaps future programs could benefit from the use of more pictures and avatars. Another option is to compare written messages with video or audio messages, so that the adherence to, and effectiveness of, the program can be improved.

Acknowledgments

The *Alerta Alcohol. Design and evaluation of a Web-based computer-tailored intervention for selective prevention of binge drinking in adolescents* project is financed by a grant from the Consejería de salud de la Junta de Andalucía, Spain (contract: PI-0031–2014), and was registered on August 4, 2015. In addition, this project was also partially financed by a grant from the Consejería de salud de la Junta de Andalucía, Spain, with the project *Activos para la Salud Positiva en la Adolescencia: Intervención Familiar Basada en Nuevas Tecnologías-Web para la Prevención del Consumo Episódico Excesivo de Alcohol* (contract: PI-0012–2017), which was registered on December 22, 2017.

Authors' Contributions

MLS is the principal investigator, developed the general concept of the study and the objectives, participated in the design of the project, and contributed to the development of the intervention, data interpretation, and the successive revisions of the paper. JMMM participated in the design of the project; contributed to the development of the intervention during the field work, development the analysis of the intervention, and interpretation of the data; and drafted the first version of paper. LM also contributed to the development of the intervention, development of the analysis of the intervention, the interpretation of the data, and the successive revisions of the paper. MC participated in the development of the analysis plan and assisted in the statistical analysis and in interpreting the results of the analysis. HdV and JSLR participated in the design of the project and in the final revision and contributed intellectually. All authors reviewed drafts of the article and gave their approval for the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2329 KB - jmir_v22i1e15438_app1.pdf]

Multimedia Appendix 2
Measurement variables from the Alerta Alcohol program.
[DOCX File , 18 KB - jmir_v22i1e15438_app2.docx]

References

1. Management of Substance Abuse Unit (MSB) in the Department of Mental Health and Substance Abuse (MSD) of the World Health Organization (WHO). Global Status Report on Alcohol and Health. Geneva, Switzerland: World Health Organization; 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639-eng.pdf?ua=1 [accessed 2020-01-09]

2. Chestnov O. Global Status Report on Alcohol and Health. Geneva, Switzerland: World Health Organization; 2014. Foreword URL: https://www.who.int/substance_abuse/publications/global_alcohol_report/msb_gsr_2014_1.pdf?ua=1 [accessed 2020-01-09]
3. Falcón M, Navarro-Zaragoza J, García-Rodríguez RM, Nova-López D, González-Navarro MD, Mercadal M, et al. Perceived barriers to implementing screening and brief intervention for alcohol consumption by adolescents in hospital emergency department in Spain. Adicciones 2018 Jul 12;30(3):189-196 [FREE Full text] [doi: 10.20882/adicciones.896] [Medline: 28749525]
4. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Sep 22;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]
5. Alcohol Facts and Statistics. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism (NIAAA); 2017. URL: https://www.niaaa.nih.gov/sites/default/files/AlcoholFactsandStats.pdf [accessed 2020-01-09]
6. Golpe S, Isorna M, Barreiro C, Braña T, Rial A. Binge drinking among adolescents: Prevalence, risk practices and related variables. Adicciones 2017 Sep 29;29(4):256-267 [FREE Full text] [doi: 10.20882/adicciones.932] [Medline: 28170056]
7. Chung T, Creswell KG, Bachrach R, Clark DB, Martin CS. Adolescent binge drinking. Alcohol Res 2018;39(1):5-15 [FREE Full text] [Medline: 30557142]
8. National Drug Plan. ESTUDES 2016/2017: Survey on Drug Use Among Secondary School Students. Madrid, Spain: Ministry of Health, Social Services and Equality; 2018 May 25. URL: http://www.pnsd.mscbs.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/2016_2017_ESTUDES.pdf [accessed 2020-01-09]
9. Mongan D, Long J. Standard Drink Measures in Europe: Peoples' Understanding of Standard Drinks and Their Use in Drinking Guidelines, Alcohol Surveys and Labelling. Dublin, Ireland: Reducing Alcohol Related Harm (RARHA); 2015 May 22. URL: https://www.researchgate.net/profile/Deirdre_Mongan/publication/322273447_Standard_drink_measures_throughout_Europe_peoples_understanding_of_standard_drinks_and_their_use_in_drinking_guidelines_alcohol_survey_and_labelling/links/5a4f78f00f7e9bbfacfd4adc/Standard-drink-measures-throughout-Europe-peoples-understanding-of-standard-drinks-and-their-use-in-drinking-guidelines-alcohol-survey-and-labelling.pdf [accessed 2020-01-09]
10. Parada M, Corral M, Caamaño-Isorna F, Mota N, Crego A, Rodríguez Holguín S, et al. Definición del concepto de consumo intensivo de alcohol adolescente (binge drinking). Adicciones 2011 Mar 01;23(1):53-63. [doi: 10.20882/adicciones.167]
11. Jander A, Mercken L, Crutzen R, de Vries H. Determinants of binge drinking in a permissive environment: Focus group interviews with Dutch adolescents and parents. BMC Public Health 2013 Sep 24;13:882 [FREE Full text] [doi: 10.1186/1471-2458-13-882] [Medline: 24063544]
12. Kaul P, Coupey SM. Clinical evaluation of substance abuse. Pediatr Rev 2002 Mar;23(3):85-94. [doi: 10.1542/pir.23-3-85] [Medline: 11875181]
13. Martínez-Hernáez Á, Marí-Klose M, Julià A, Escapa S, Marí-Klose P. Consumo episódico excesivo de alcohol en adolescentes: su asociación con los estados de ánimo negativos y los factores familiares. Revista Española de Salud Pública 2012 Feb;86(1):101-114 [FREE Full text] [doi: 10.1590/s1135-57272012000100009]
14. Hunter Z, Francescutti LH. Facing the consequences of binge drinking. Can Fam Physician 2013 Oct;59(10):1041-1042, e438 [FREE Full text] [Medline: 24130272]
15. Anderson P. Binge Drinking and Europe. London, UK: Institute of Alcohol Studies; 2007 Jul. URL: http://www.drugsandalcohol.ie/6353/1/3836-4088.pdf [accessed 2018-10-01]
16. Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a Web-based computer-tailored game to reduce binge drinking among Dutch adolescents: A cluster randomized controlled trial. J Med Internet Res 2016 Feb 03;18(2):e29 [FREE Full text] [doi: 10.2196/jmir.4708] [Medline: 26842694]
17. Best D, Manning V, Gossop M, Gross S, Strang J. Excessive drinking and other problem behaviours among 14-16 year old schoolchildren. Addict Behav 2006 Aug;31(8):1424-1435. [doi: 10.1016/j.addbeh.2005.12.002] [Medline: 16442742]
18. Cox W, Fadardi J, Hosier S, Pothos E. Differential effects and temporal course of attentional and motivational training on excessive drinking. Exp Clin Psychopharmacol 2015 Dec;23(6):445-454 [FREE Full text] [doi: 10.1037/pha0000038] [Medline: 26348159]
19. Wechsler H, Dowdall GW, Maenner G, Gledhill-Hoyt J, Lee H. Changes in binge drinking and related problems among American college students between 1993 and 1997: Results of the Harvard School of Public Health College Alcohol Study. J Am Coll Health 1998 Sep;47(2):57-68. [doi: 10.1080/07448489809595621] [Medline: 9782661]
20. O'Malley P, Johnston L, Bachman J. Alcohol use among adolescents. Alcohol Health Res World 1998;22(2):85-93 [FREE Full text] [Medline: 15706782]
21. ESPAD Group. ESPAD Report 2015: Results From the European School Survey Project on Alcohol and Other Drugs. Luxembourg: Publications Office of the European Union; 2016. URL: http://www.espad.org/sites/espad.org/files/TD0116475ENN.pdf [accessed 2020-01-09]
22. Newton N, Andrews G, Teesson M, Vogl L. Delivering prevention for alcohol and cannabis using the Internet: A cluster randomised controlled trial. Prev Med 2009 Jun;48(6):579-584. [doi: 10.1016/j.ypmed.2009.04.009] [Medline: 19389420]

23. Krebs P, Prochaska J, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
24. Velicer WF, Redding CA, Paiva AL, Mauriello LM, Blissmer B, Oatley K, et al. Multiple behavior interventions to prevent substance abuse and increase energy balance behaviors in middle school students. Transl Behav Med 2013 Mar;3(1):82-93 [FREE Full text] [doi: 10.1007/s13142-013-0197-0] [Medline: 23585821]
25. Drost R, Paulus A, Jander A, Mercken L, de Vries H, Ruwaard D, et al. A Web-based computer-tailored alcohol prevention program for adolescents: Cost-effectiveness and intersectoral costs and benefits. J Med Internet Res 2016 Apr 21;18(4):e93 [FREE Full text] [doi: 10.2196/jmir.5223] [Medline: 27103154]
26. Brick L, Redding C, Paiva A, Velicer W. Intervention effects on stage transitions for adolescent smoking and alcohol use acquisition. Psychol Addict Behav 2017 Aug;31(5):614-624. [doi: 10.1037/adb0000302] [Medline: 28714725]
27. Schulz D, Candel M, Kremers S, Reinwand D, Jander A, de Vries H. Effects of a Web-based tailored intervention to reduce alcohol consumption in adults: Randomized controlled trial. J Med Internet Res 2013 Sep 17;15(9):e206 [FREE Full text] [doi: 10.2196/jmir.2568] [Medline: 24045005]
28. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: Introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [doi: 10.1016/s0738-3991(98)00127-x] [Medline: 10223015]
29. Lustria M, Noar S, Cortese J, Van Stee SK, Glueckauf R, Lee J. A meta-analysis of Web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
30. Walton M, Chermack S, Shope J, Bingham C, Zimmerman M, Blow F, et al. Effects of a brief intervention for reducing violence and alcohol misuse among adolescents: A randomized controlled trial. JAMA 2010 Aug 04;304(5):527-535 [FREE Full text] [doi: 10.1001/jama.2010.1066] [Medline: 20682932]
31. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub M. Efficacy of a Web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Feb;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
32. De Bourdeaudhuij I, Maes L, De Henauw S, De Vriendt T, Moreno LA, Kersting M, HELENA Study Group. Evaluation of a computer-tailored physical activity intervention in adolescents in six European countries: The Activ-O-Meter in the HELENA intervention study. J Adolesc Health 2010 May;46(5):458-466. [doi: 10.1016/j.jadohealth.2009.10.006] [Medline: 20413082]
33. Schuckit M, Smith T, Kalmijn J, Skidmore J, Clausen P, Shafir A, et al. The impact of focusing a program to prevent heavier drinking on a pre-existing phenotype, the low level of response to alcohol. Alcohol Clin Exp Res 2015 Feb;39(2):308-316 [FREE Full text] [doi: 10.1111/acer.12620] [Medline: 25656349]
34. Elfeddali I, Bolman C, Candel M, Wiers RW, de Vries H. Preventing smoking relapse via Web-based computer-tailored feedback: A randomized controlled trial. J Med Internet Res 2012 Aug 20;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]
35. de Vries H, Logister M, Krekels G, Klaasse F, Servranckx V, van Osch L. Internet based computer tailored feedback on sunscreen use. J Med Internet Res 2012 Apr 30;14(2):e48 [FREE Full text] [doi: 10.2196/jmir.1902] [Medline: 22547528]
36. Kelders S, Kok R, Ossebaard H, 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]
37. 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]
38. Jander A, Crutzen R, Mercken L, de Vries H. A Web-based computer-tailored game to reduce binge drinking among 16 to 18 year old Dutch adolescents: Development and study protocol. BMC Public Health 2014 Oct 09;14:1054 [FREE Full text] [doi: 10.1186/1471-2458-14-1054] [Medline: 25301695]
39. Lima-Serrano M, Martínez-Montilla JM, Lima-Rodríguez JS, Mercken L, de Vries H. Design, implementation and evaluation of a Web-based computer-tailored intervention to prevent binge drinking in adolescents: Study protocol. BMC Public Health 2018 Apr 04;18(1):449 [FREE Full text] [doi: 10.1186/s12889-018-5346-4] [Medline: 29618331]
40. Pearson M, Chilton R, Wyatt K, Abraham C, Ford T, Woods H, et al. Implementing health promotion programmes in schools: A realist systematic review of research and experience in the United Kingdom. Implement Sci 2015 Oct 28;10:149 [FREE Full text] [doi: 10.1186/s13012-015-0338-6] [Medline: 26510493]
41. Moher D, Schulz KF, Altman DG, CONSORT. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 2001;1:2 [FREE Full text] [doi: 10.1186/1471-2288-1-2] [Medline: 11336663]
42. Martinez-Montilla JM, Lima-Serrano M, Mercken L, de Vries H, Lima-Rodríguez JS. Alerta Alcohol. 2020. URL: http://institucional.us.es/alertalcohol/ [accessed 2020-01-02]
43. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New York, NY: Prentice Hall; 1986.
44. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): An example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626. [doi: 10.1093/her/cyg031] [Medline: 14572020]

45. Nugawela M, Langley T, Szatkowski L, Lewis S. Measuring alcohol consumption in population surveys: A review of international guidelines and comparison with surveys in England. Alcohol Alcohol 2016 Jan;51(1):84-92 [FREE Full text] [doi: 10.1093/alcalc/agv073] [Medline: 26115987]
46. Lima-Serrano M, Vargas-Martínez AM, Gil García E, Martínez-Montilla JM, Lima Rodríguez JS, de Vries H. Adaptation and validation of five scales to measure determinants of binge drinking in adolescents [Article in Spanish]. An Sist Sanit Navar 2017 Aug 31;40(2):221-236 [FREE Full text] [doi: 10.23938/ASSN.0028] [Medline: 28765653]
47. Currie C, Molcho M, Boyce W, Holstein B, Torsheim T, Richter M. Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. Soc Sci Med 2008 Mar;66(6):1429-1436. [doi: 10.1016/j.socscimed.2007.11.024] [Medline: 18179852]
48. Smilkstein G, Ashworth C, Montano D. Validity and reliability of the family APGAR as a test of family function. J Fam Pract 1982 Aug;15(2):303-311. [Medline: 7097168]
49. Bannink R, Broeren S, Joosten-van Zwanenburg E, van As E, van de Looij-Jansen P, Raat H. Effectiveness of a Web-based tailored intervention (E-health4Uth) and consultation to promote adolescents' health: Randomized controlled trial. J Med Internet Res 2014 May 30;16(5):e143 [FREE Full text] [doi: 10.2196/jmir.3163] [Medline: 24878521]
50. National Drug Plan. ESTUDES 2014/2015: Survey on Drug Use Among Secondary School Students. Madrid, Spain: Ministry of Health, Social Services and Equality; 2016 Feb 11. URL: http://www.pnsd.mscbs.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/2016_ESTUDES_2014-2015.pdf [accessed 2020-01-12]
51. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
52. White I, Carlin J. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med 2010 Dec 10;29(28):2920-2931. [doi: 10.1002/sim.3944] [Medline: 20842622]
53. White I, Royston P, Wood A. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
54. Measham F. The 'Big Bang' approach to sessional drinking: Changing patterns of alcohol consumption amongst young people in North West England. Addict Res 2009 Jul 11;4(3):283-299. [doi: 10.3109/16066359609005573]
55. Romo-Avilés N, Marcos-Marcos J, Marquina-Márquez A, Gil-García E. Intensive alcohol consumption by adolescents in Southern Spain: The importance of friendship. Int J Drug Policy 2016 May;31:138-146. [doi: 10.1016/j.drugpo.2016.01.014] [Medline: 26948500]
56. Kuntsche E, Knibbe R, Gmel G, Engels R. Why do young people drink? A review of drinking motives. Clin Psychol Rev 2005 Nov;25(7):841-861. [doi: 10.1016/j.cpr.2005.06.002] [Medline: 16095785]
57. Donath C, Grässel E, Baier D, Pfeiffer C, Karagülle D, Bleich S, et al. Alcohol consumption and binge drinking in adolescents: Comparison of different migration backgrounds and rural vs urban residence--A representative study. BMC Public Health 2011 Feb 07;11:84 [FREE Full text] [doi: 10.1186/1471-2458-11-84] [Medline: 21299841]
58. de Ruijter D, Smit E, de Vries H, Hoving C. Web-based computer-tailoring for practice nurses aimed to improve smoking cessation guideline adherence: A study protocol for a randomized controlled effectiveness trial. Contemp Clin Trials 2016 May;48:125-132. [doi: 10.1016/j.cct.2016.04.007] [Medline: 27103232]
59. Stanczyk N, de Vries H, Candel M, Muris J, Bolman CA. Effectiveness of video- versus text-based computer-tailored smoking cessation interventions among smokers after one year. Prev Med 2016 Jan;82:42-50 [FREE Full text] [doi: 10.1016/j.ypmed.2015.11.002] [Medline: 26577867]
60. Schulz D, Kremers S, Vandelanotte C, van Adrichem MJ, Schneider F, Candel M, et al. Effects of a Web-based tailored multiple-lifestyle intervention for adults: A two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014 Jan 27;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
61. de Vries H, Kremers S, Smeets T, Brug J, Eijmael K. The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors. Am J Health Promot 2008;22(6):417-425. [doi: 10.4278/ajhp.22.6.417] [Medline: 18677882]
62. Prochaska JO, Redding CA, Evers KE. The Transtheoretical Model and stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. 4th edition. San Francisco, CA: Jossey-Bass; 2008:97-121.
63. Schneider F, de Vries H, Candel M, van de Kar A, van Osch L. Periodic email prompts to re-use an Internet-delivered computer-tailored lifestyle program: Influence of prompt content and timing. J Med Internet Res 2013 Jan 31;15(1):e23 [FREE Full text] [doi: 10.2196/jmir.2151] [Medline: 23363466]
64. Cremers H, Mercken L, Oenema A, de Vries H. A Web-based computer-tailored smoking prevention programme for primary school children: Intervention design and study protocol. BMC Public Health 2012 Jun 11;12:277 [FREE Full text] [doi: 10.1186/1471-2458-12-277] [Medline: 22490110]
65. Schulz D, Schneider F, de Vries H, van Osch LA, van Nierop PW, Kremers SP. Program completion of a Web-based tailored lifestyle intervention for adults: Differences between a sequential and a simultaneous approach. J Med Internet Res 2012 Mar 08;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]

66. Schneider F, van Osch L, Schulz D, Kremers S, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an Internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012 Mar 01;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
67. Kohl L, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: A systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]

Abbreviations

BD: binge drinking
CC: control condition
CONSORT: Consolidated Standards of Reporting Trials
CSE: compulsory secondary education
CT: computer tailored
EC: experimental condition
ESPAD: European School Survey Project on Alcohol and Other Drugs
ESTUDES: Encuesta Sobre Uso de Drogas en Estudiantes de Enseñanzas Secundarias
FAS: Family Affluence Scale
GDPR: General Data Protection Regulation
HED: heavy episodic drinking
OR: odds ratio
SDU: standard drink unit
VT: vocational training

Edited by G Eysenbach; submitted 10.07.19; peer-reviewed by A Videira-Silva, P Wallace, A Al-Zalabani; comments to author 13.08.19; revised version received 04.10.19; accepted 19.10.19; published 24.01.20.

Please cite as:
Martinez-Montilla JM, Mercken L, de Vries H, Candel M, Lima-Rodríguez JS, Lima-Serrano M
A Web-Based, Computer-Tailored Intervention to Reduce Alcohol Consumption and Binge Drinking Among Spanish Adolescents: Cluster Randomized Controlled Trial
J Med Internet Res 2020;22(1):e15438
URL: http://www.jmir.org/2020/1/e15438/
doi:10.2196/15438
PMID:32012064

Original Paper

An Interactive Web-Based Lethal Means Safety Decision Aid for Suicidal Adults (Lock to Live): Pilot Randomized Controlled Trial

Marian E Betz[1,2], MPH, MD; Christopher E Knoepke[3,4], MSW, PhD; Scott Simpson[5,6], MD, MPH; Bonnie J Siry[1], MSS; Ashley Clement[1], MSW; Tamara Saunders[7], BA; Rachel Johnson[8], MS; Deborah Azrael[9], PhD; Edwin D Boudreaux[10], PhD; Faris Omeragic[1], BS; Leah M Adams[1], MPH; Sydney Almond[11]; Elizabeth Juarez-Colunga[4,8], PhD; Daniel D Matlock[2,4,12], MPH, MD

[1]Department of Emergency Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[2]Eastern Colorado Geriatric Research Education and Clinical Center, Veterans Health Administration, Aurora, CO, United States

[3]Division of Cardiology, University of Colorado School of Medicine, Aurora, CO, United States

[4]Adult & Child Consortium for Outcomes Research & Delivery Science, University of Colorado School of Medicine, Aurora, CO, United States

[5]Psychiatric Emergency Services, Denver Health Medical Center, Denver, CO, United States

[6]Department of Psychiatry, University of Colorado School of Medicine, Aurora, CO, United States

[7]School of Public Affairs, University of Colorado Colorado Springs, Colorado Springs, CO, United States

[8]Department of Biostatistics & Informatics, Colorado School of Public Health, Aurora, CO, United States

[9]Harvard Injury Control Research Center, Harvard School of Public Health, Boston, MA, United States

[10]Departments of Emergency Medicine, Psychiatry, and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[11]University of Colorado Denver, Denver, CO, United States

[12]Division of Geriatric Medicine, Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United States

Corresponding Author:
Marian E Betz, MPH, MD
Department of Emergency Medicine
University of Colorado School of Medicine
12401 E 17th Avenue
B-215
Aurora, CO, 80045
United States
Phone: 1 7208486777
Email: marian.betz@cuanschutz.edu

Abstract

Background: Counseling to reduce access to lethal means such as firearms and medications is recommended for suicidal adults but does not routinely occur. We developed the Web-based *Lock to Live* (L2L) decision aid to help suicidal adults and their families choose options for safer home storage.

Objective: This study aimed to test the feasibility and acceptability of L2L among suicidal adults in emergency departments (EDs).

Methods: At 4 EDs, we enrolled participants (English-speaking, community-dwelling, suicidal adults) in a pilot randomized controlled trial. Participants were randomized in a 13:7 ratio to L2L or control (website with general suicide prevention information) groups and received a 1-week follow-up telephone call.

Results: Baseline characteristics were similar between the intervention (n=33) and control (n=16) groups. At baseline, many participants reported having access to firearms (33/49, 67%), medications (46/49, 94%), or both (29/49, 59%). Participants viewed L2L for a median of 6 min (IQR 4-10 min). L2L also had very high acceptability; almost all participants reported that they would recommend it to someone in the same situation, that the options felt realistic, and that L2L was respectful of values about firearms. In an exploratory analysis of this pilot trial, more participants in the L2L group reported reduced firearm access at follow-up, although the differences were not statistically significant.

Conclusions: The L2L decision aid appears feasible and acceptable for use among adults with suicide risk and may be a useful adjunct to lethal means counseling and other suicide prevention interventions. Future large-scale studies are needed to determine the effect on home access to lethal means.

Trial Registration: ClinicalTrials.gov NCT03478501; https://clinicaltrials.gov/ct2/show/NCT03478501

(J Med Internet Res 2020;22(1):e16253) doi:10.2196/16253

KEYWORDS

internet; firearm; suicide; medication

Introduction

Identification and intervention with adults at risk of suicide are recommended for health care settings, including emergency departments (EDs) and primary care settings [1]. Reducing access to lethal means for those at risk of suicide is an evidence-based suicide prevention strategy [2,3]. As such, lethal means counseling (LMC) by health care providers is recommended by multiple professional organizations and is a goal of the National Strategy for Suicide Prevention [3].

Firearms are a particular focus for LMC because of their high case fatality rate (85-90%, much higher than other methods) and widespread availability [4,5]. Medical [6,7], suicide prevention [8], and firearm organizations [9] have advocated for securing firearms or removing them from the homes of persons at risk of suicide. Medications are a second key focus of LMC, as overdose is the leading cause of nonfatal suicide attempts [10] and death rates because of opioid overdose are increasing. In addition, including medications in LMC may help improve the acceptability of conversations about firearms [11,12].

Previous work suggests that ED clinicians bring up firearm safety with fewer than half of suicidal patients [13,14]. Likely barriers to counseling include inadequate provider training and awareness along with time demands on busy clinicians; clinicians may also be uncomfortable bringing up a sensitive topic, although previous work shows that patients are generally open to respectful discussion [11,15-17]. Yet the ED is a key acute care setting for suicide prevention, as it is typically the location where patients with suicidal ideation or a suspected suicide attempt are referred for evaluation.

To address these constraints, our team developed the Web-based *Lock to Live* (L2L) decision aid [18] for suicidal adults and their families to consider "which options to choose to reduce home access to firearms or medications" (Multimedia Appendix 1). The L2L Web-based decision aid was developed through an iterative process based on qualitative interviews with key stakeholders, including suicide prevention experts, members of the firearm community, survivors of suicide attempts, and loved ones of suicide victims, as described elsewhere [19]. It includes the typical decision aid sections to help an individual understand a decision (eg, identifying personal preferences and exploring options) and make a choice. Specifically, L2L walks an individual through the rationale for reducing home firearm or medication access and then explores preferences (such as cost) and in- and out-of-home storage options. We hypothesized that a self-administered decision aid to engage ED patients in decision making and augment routine counseling could ultimately enhance patient outcomes, provider satisfaction, and ease of implementation and dissemination. An electronic, Web-based format might facilitate implementation by avoiding the need for paper forms in clinical settings and by allowing confidential engagement for patients waiting in clinical settings; a Web-based format could also allow for broader dissemination to other settings (eg, at home or in outpatient settings).

Here, we describe the results of a pilot randomized controlled trial in EDs that aimed to test the feasibility and acceptability of L2L for adults with suicidal ideation or behavior. Results from this pilot trial can inform the implementation of LMC for adults in EDs and testing of L2L in non-ED settings.

Methods

Design and Recruitment

This pilot feasibility trial recruited participants from 4 large EDs in Colorado: a tertiary care academic center, an urban safety net hospital, and a regional medical center with 2 EDs in a geographic region with firearm ownership rates that are higher than state averages. All EDs had 24/7 coverage by behavioral health specialists. Study procedures occurred in the area where the patient was receiving clinical care to limit disruption to ED care and maintain safety precautions. As this was a multisite trial involving a vulnerable population, this study was approved through a full board review of the Colorado Multiple Instructional Review Board, and it was monitored by an external data safety and monitoring board. The trial was registered at Clinicaltrials.gov (NCT03478501).

Eligible participants were English-speaking adult (≥18 years) patients identified as having suicide risk who were not in police custody, who were willing and able to complete telephone follow-up at 1 week (eg, had a working telephone), and who reported ≥1 firearm at home. The study was later expanded to patients with any medications at home (see below). Potentially eligible patients were identified by research assistants (RAs) and approached once deemed medically stable and sober by the treating ED team. Other psychiatric complaints or symptoms (eg, hallucinations) did not preclude eligibility screening, although research staff used discretion in approaching agitated or violent patients. The consent process included questions to determine cognitive capacity to consent. The eligibility and consent script guided the RA to establish rapport, discuss the larger goal of the study (improving home safety generally), and explain participation and confidentiality before asking about firearm ownership.

At each site, participants were a priori block randomized preferentially to the intervention group (13:7 ratio) to increase the amount of feedback on L2L. Randomization occurred after consent to minimize enrollment bias. Participants were blinded but the research staff were not. To blind participants, we used mild deception in the informed consent process, such that patients knew the study was examining ways to enhance home safety of suicidal patients but did not know that L2L was the intervention of interest. The clinical staff were unaware of the treatment group.

At enrollment, participants completed a Web-based baseline questionnaire and then viewed either (1) L2L on a Web-enabled tablet computer or (2) the control, also on the tablet, consisting of general suicide prevention information without a focus on firearm or medication storage. All participants then completed a second questionnaire, including indicating their plan for firearm and medication storage and acceptability questions for the intervention group. Participants could choose to receive a paper printout of either their chosen storage options (L2L group) or the control information. Participating patients were contacted by telephone 1 week after the ED visit for a short questionnaire, which included questions about current firearm and medication storage and changes since the ED visit. Participants who did not answer the telephone for the 1-week follow-up phone call were contacted about once a week thereafter. Participants were considered lost to follow-up at 3 months after enrollment. On the basis of the medical records and vital statistics review, we confirmed that no participants (either contacted or lost to follow-up) had died.

On the basis of the initial feedback and low recruitment, and with institutional review board and data safety monitoring board approval, the study team modified L2L and the study protocol partway through the trial to also address medication safety in addition to firearms. Specifically, L2L incorporated a module on reducing access to medications, and patients could be eligible if they reported medications at home (even without firearms); other eligibility criteria and study procedures were unchanged. The medication module was developed following the same methods used for original L2L development, including stakeholder interviews and iterative refinement [19].

Measurements

Study data were collected and managed using Research Electronic Data Capture (REDCap), a secure Web system. Participants self-completed questionnaires on tablet computers, and RAs entered additional data (eg, time required to complete L2L) afterward. REDCap was also used for baseline medical record abstraction and telephone follow-up questionnaires (administered by RA).

Key measures assessed in the intervention group included feasibility and acceptability. Feasibility was measured via minutes for the patient to complete L2L as measured by research staff, along with completion rate. Acceptability was measured using the Ottawa Acceptability Scale, a scale measuring comprehensibility (eg, length, amount of information, balance in presentation, and overall suitability for decision making) [20].

Although the pilot trial was not powered to measure the efficacy of L2L on decisions or behavior, for exploratory analysis, we measured (1) decision conflict, a fundamental component of decision quality as a precursor to behavior change [21], and (2) behavior change itself. We hypothesized that patients with higher quality decisions (defined as lower decision conflict) after L2L would be more likely to change their home storage to reduce access to lethal means. Decision conflict was measured using the low-literacy version of the Decisional Conflict Scale (DCS), a 10-item scale with high reliability and test-retest correlation previously shown to discriminate between known groups who make or delay decisions [22]. The DCS scale is scored from 0 to 100, with lower scores indicating less decisional conflict. The baseline and follow-up questionnaires also recorded demographics, living situation, home firearms and medications, and suicide ideation or attempts as measured by the baseline and since-last-visit versions of the Columbia-Suicide Severity Rating Scale [23]. For behavior change, we examined changes in firearm or medication storage between enrollment and follow-up, categorized as *moved in safer direction* (eg, using new or more locking devices or moving items out of the home), *no change*, or *moved in less safe direction* (eg, use of fewer locking devices).

Statistical Analysis

We used descriptive statistics for feasibility, acceptability, and exploratory analyses on DCS and behavior change. For continuous variables, differences in means between control and intervention groups were tested with 2-sample *t* tests with unequal variances. For categorical variables, we used frequencies with percentages, and differences between groups were tested with Fisher exact test.

Results

Participant Characteristics

Over 10 months, 49 patients were enrolled, with 33 randomized to the L2L intervention group and 16 to the control group (Figure 1). Overall, patient participant characteristics and study results were similar before and after expanding L2L and eligibility criteria, so results are presented in aggregate. Intervention and control groups were similar on key characteristics and measures (Table 1).

At baseline, 33 of 49 (67%) participants reported having access to firearms at home, 94% (46/49) had medications at home, and 59% (29/49) had both. Of the 33 patients with at least one known firearm at home, 11 (33%) reported 1, 19 (58%) reported more than 1, and 3 (9%) were not sure how many there were. When asked about baseline storage, many of these suicidal adults reported that at least one firearm was unlocked (15/33, 45%), loaded (12/33, 36%), or both unlocked and loaded (8/33, 24%). For the 27 participants with locked firearms, locking devices (eg, cable or trigger locks) were most common (12/27, 44%), followed by firearm safes (9/27, 33%). Of the 47 patients with medications at home, participants reported storing unlocked some or all prescription pain medications (27/47, 57%), over-the-counter pain medications (29/47, 62%), other prescription medications (64%, n=30), and other over-the-counter medications (33/47, 70%). More than half of

the 47 enrolled patients with medications at home reported that all were stored unlocked (33/47, 70%).

On the basis of the medical record review, most participants (43/49, 88%) were evaluated by a mental health professional in the ED, and most participants (41/49, 84%) had documentation that at least one provider assessed their access to lethal means. When asked about their ED care, more participants remembered at least one provider talking to them about home firearm access (32/49, 65%) than about medication storage (11/49, 22%).

Figure 1. Consolidated Standards of Reporting Trials diagram.

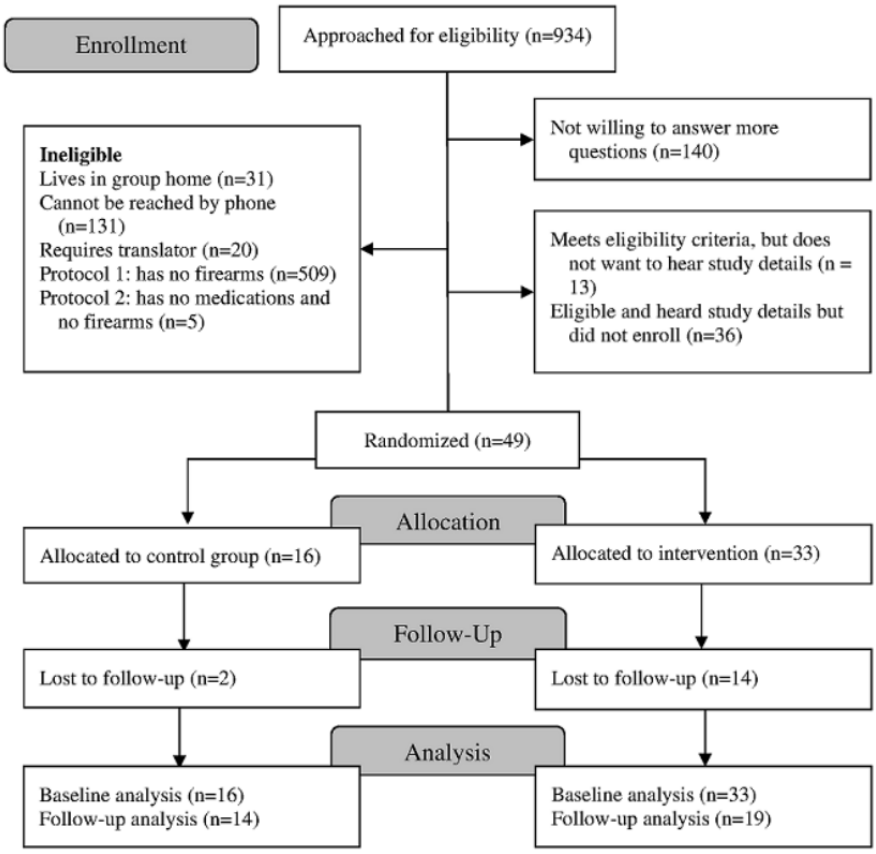


Table 1. Participant characteristics by study group (enrollment).

Participant characteristic[a]	*Lock to Live* (N=33)	Control (N=16)
Age (years), mean (range)	39 (21-68)	38 (21-69)
Male, n (%)	17 (52)	9 (56)
Veteran, n (%)	6 (18)	5 (31)
Currently employed, n (%)	16 (48)	9 (56)
≥1 child at home, n (%)	8 (24)	6 (38)
Currently receiving outpatient mental health care, n (%)	7 (21)	7 (44)
Race, n (%)		
White	23 (70)	10 (63)
Black	5 (15)	0 (0)
Other or not documented	5 (15)	6 (38)
Hispanic ethnicity	5 (15)	4 (25)
Highest grade completed, n (%)		
High school graduate or less	9 (27)	4 (25)
Vocational or technical school graduate or some college	14 (42)	10 (63)
College graduate or higher	10 (30)	2 (13)
Current marital status, n (%)		
Never married	15 (45)	4 (25)
Married	10 (30)	8 (50)
Widowed, divorced, or other	8 (24)	4 (25)
Columbia-Suicide Severity Rating Scale, n (%)		
Lifetime	12.9 (6)	14.4 (5.7)
Past week	16.4 (4)	14.8 (5.9)
Medical record documentation, n (%)		
Alcohol abuse	4 (12)	5 (31)
Alcohol intoxication	6 (18)	5 (31)
Intentional illegal or prescription drug use	11 (33)	3 (19)
Past week (including current visit), n (%)		
Suicide ideation	31 (94)	13 (81)
Suicide attempt	12 (36)	6 (38)
Homicidal ideation	3 (9)	1 (6)
Interpersonal violence	3 (9)	1 (6)
Lethal means access, n (%)		
Not documented	5 (15)	3 (19)
No access to lethal means	4 (12)	1 (6)
Access to lethal means	20 (61)	10 (63)
Firearms specifically mentioned	18 (90)	10 (100)
Way to reduce access to lethal means discussed, n (%)		
Yes, documented in medical record	9 (27)	9 (56)
No, documented in medical record	12 (36)	2 (13)
Not documented in medical record	9 (27)	3 (19)
Evaluated by mental health professional during visit, n (%)	29 (88)	14 (88)
Written safety plan, n (%)	3 (9)	0 (0.0)

Participant characteristic[a]	*Lock to Live* (N=33)	Control (N=16)
Decision outcomes, mean (SD)		
Total DCS[b] score (out of 100)	12.6 (20)	9.7 (18)
DCS score <25[c]	29 (88)	13 (81)
Likelihood of changing storage at home (out of 7), mean (SD)		
Firearms	3.5 (2.5)	4.2 (2.4)
Medications	3.6 (2.3)	4.1 (2.5)

[a]Numbers and percentages may not sum to total because of missing data (not shown if <5%) or questions allowing ≥1 response.

[b]DCS: Decisional Conflict Scale.

[c]Scores <25 previously associated with implementing decisions [22].

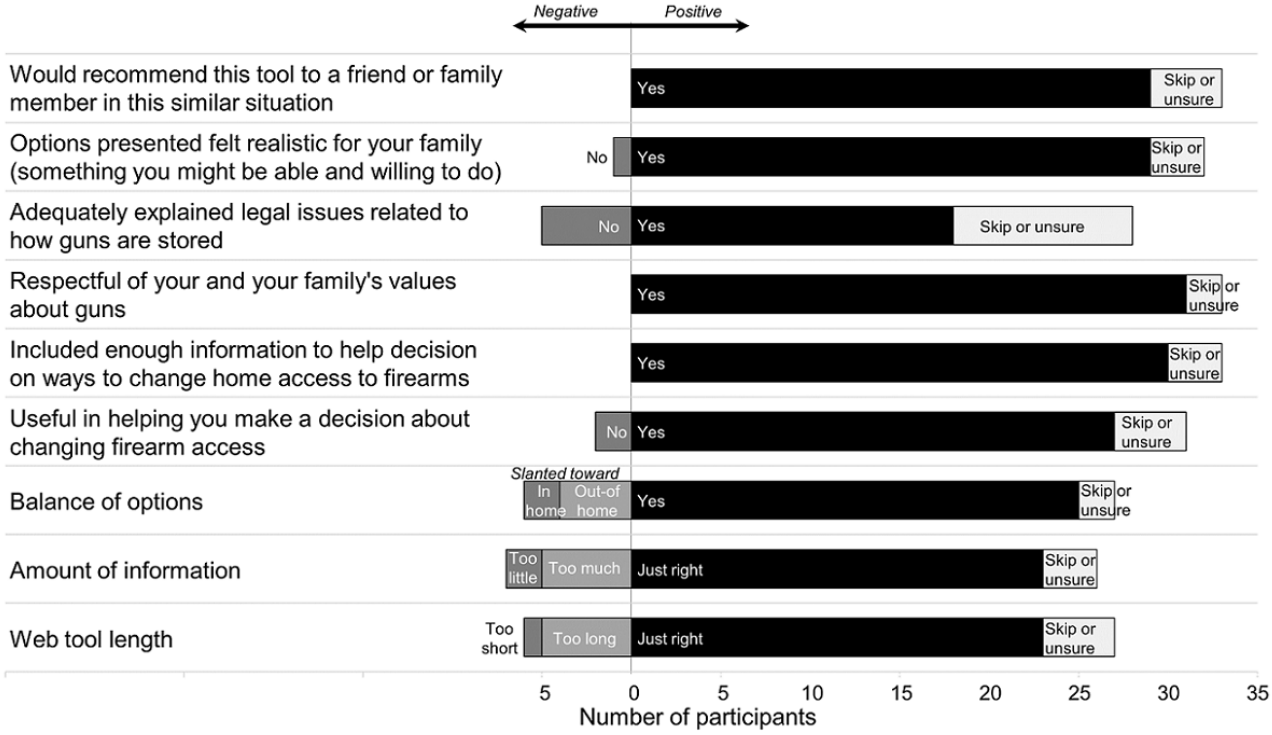
Intervention Feasibility and Acceptability

Feasibility and acceptability of the L2L intervention were excellent. All intervention group participants (n=33) completed L2L, with a median viewing time of 6 min (IQR 4-10 min), and most (24/33, 73%) of the participants wanted a printout of the last page with the final choices and recommendations. Most participants (31/33, 94%) viewed L2L by themselves, without a family or friend present and without a provider. Multimedia Appendix 2 displays selected storage options; participants made 53 selections in L2L, in addition to *friends or family* (which was selected by default but could be unselected; 10 participants unselected it). Figure 2 shows responses to the Ottawa Acceptability Scale. Almost all participants reported that they would recommend the tool to a friend or family member in the same situation, that the options felt realistic, and that L2L was respectful of values about firearms. Areas for improvement included explanation of legal issues, as 15% (5/33) of the participants reported that the tool did not adequately explain legal issues and 30% (10/33) of the participants were unsure or skipped the question. Most of the participants felt that the tool had the right length and had the right amount of information, with a balance in the presentation of options.

In the exploratory analysis, there were no significant differences in decision conflict or planned storage changes between the L2L and control groups. DCS scores were low in both groups (suggesting low decision conflict), and participants gave overall neutral responses when asked about the likelihood of changing either firearm or medication storage (mean 3.8 out of 7). When asked about planned changes to storage, 8 participants overall said that they would use more lockboxes at home or dispose of unneeded medications (8/31, 26% each), with no difference between the L2L and control groups. For firearms, participants across the L2L and control groups most commonly said that they would use more firearm locking devices (14/22, 64% overall) or safes or lockboxes (13/22, 59% overall) at home; the most common out-of-home options were storing firearms with a trusted family member (13/22, 59% overall) or friend (9/22, 41% overall). Fewer participants cited storage with firearm stores or ranges or with law enforcement (3/22, 14% for each) as likely.

Two-thirds (n=33; n=14 control and n=19 intervention) of the participants completed telephone follow-up (Figure 1) at an average of 2.4 weeks (SD 2.2; range 1-9 weeks) after enrollment; among these, 25 participants reported on firearm storage and 6 on medication storage. There were 14 participants who reported having firearms at both baseline and follow-up (including 3 who had moved firearms out of the home). Among these 14, as compared with the control group, more participants in the L2L group had moved in a safer direction (1/7, 14% vs 4/7, 57%), although the difference was not statistically significant. Similarly, as compared with the control group, fewer participants in the L2L group had made no change (5/7, 71% vs 3/7, 43%) or had moved in a less safe direction (1/7, 14% vs 0/7, 0%). These differences were also not statistically significant.

Figure 2. Lock to Live acceptability (n=33).



Discussion

Principal Findings

In this pilot trial, the acceptability of the L2L decision aid was very high among adults with acute suicide risk. Although the trial was not powered to identify an effect on subsequent home lethal means access, L2L appears feasible for clinical use in that it took a median of 6 min, and there were no major issues accessing the content via a tablet in the ED. Questions remain, however, about how best to implement its routine use in clinical settings for appropriate patients.

Participants overwhelmingly found L2L to be useful, informative, and respectful of their values with regard to firearm ownership. This high acceptability supports further clinical application and evaluation of L2L. Future research should address effects on decision making and subsequent lethal means access, along with methods for implementation and dissemination. It may be that L2L would have the greatest effect when used within a conversation with a provider, after rapport has already been established. In such a role, L2L might support counseling and decision making by helping an individual clarify values and understand logistics that the counselor may not be fully knowledgeable about.

On the other hand, there is evidence that patients may not disclose suicidality [24] or firearm ownership to providers, so upstream, community-based (nonclinical) approaches may also be useful to disseminate messages to those at risk. Indeed, although designed and pilot tested in ED settings, L2L does not include language specific to the ED setting and therefore could be tested in other clinical or community settings. Active studies are testing L2L in outpatient primary care and outpatient mental health settings, but integration into broader public education campaigns also deserves consideration. Such campaigns should incorporate information intended to raise awareness and change beliefs concerning the importance of reducing access to lethal means during times of suicide risk [25]—beliefs previously suggested to be associated with firearm storage behaviors [26,27].

Indeed, our pilot findings raise questions about whether suicidal adults in EDs recognize or believe that lethal means access is an issue in need of a decision or behavior change. Decision conflict scores were relatively low in both groups, with 86% (29/33) of participants overall having scores less than 25. In previous work, low decision conflict scores—indicating low internal conflict about the decision—have been associated with implementing decisions [22]. It may be that L2L would have optimal effects when integrated into care after delivery of LMC by providers, as such counseling might prime the individual for decision making. Pairing with other interventions to overcome obstacles in making storage changes—be they financial, logistical, or emotional—may also be useful. For example, in this pilot study, we did not test the effect of providing locking devices or contact information for nearby out-of-home storage locations, but such tangible add-ons may help motivate change. Previous work suggests that provision of locking devices can boost responses to LMC [28]; similarly, facilitating connection to local storage partners may overcome logistic barriers [29].

Testing the role of family or friends was a challenge in this study. There is real-world variability in how family or friends are involved with lethal means safety counseling; ideally, the person with decision-making control over storage would be involved in the ED, but they may not be present with the patient. The low observed rates of presence and participation by trusted individuals could reflect that suicidal adults may have thought distortions such that they do not want to *burden* family or friends by engaging them in research or being in the ED with them.

Low referral of family or friends may also have stemmed from confidentiality concerns or fear of firearm confiscation. Recognizing the importance of engaging trusted others, the L2L tool includes a section where patients are prompted to consider who they would enlist to help implement their storage plan (eg, family members, friends, religious leaders, or fellow Veterans). Our findings, and the questions they raise, highlight the need for strategies to better enlist patients' caring contacts in both clinical care and research participation.

Limitations

By their nature, pilot studies do not offer power to examine primary efficacy outcomes, so we designed the trial to provide information about feasibility. We cannot comment on L2L's effect on suicide-related outcomes, and we did not examine other predictors of these outcomes. The trial itself did not engage ED providers, who would be key partners in the implementation of the intervention into clinical practice; a related mixed method study is underway. Another limitation is blinding, as RAs had to know group assignment to ensure the participant could access L2L. Concerns about confidentiality expressed by stakeholders involved in the design process led to the a priori decision to not collect these data through the tool itself, however [30].

Conclusions

The line of investigation exemplified by this trial offers the possibility of better facilitating lethal means safety counseling in a patient-centered, acceptable, and feasible way. The Web format and respectful messaging offer a way for medical and mental health providers to augment LMC by providing a simple, patient-centered tool with which to present various safe storage options. For providers who are unfamiliar with options or uncomfortable with these discussions, L2L might increase their willingness to address lethal means access with at-risk patients; further examination of L2L's effect on provider behavior is warranted. Critically, if L2L is found effective in future work, its Web format offers the potential for rapid and widespread dissemination and integration into suicide prevention efforts as well as more rapid updating as indicated (eg, for new relevant legislation). Future mixed methods research examining tool effectiveness and implementation could help enhance home safety and prevent suicide.

Acknowledgments

This research was supported by the National Institutes of Health (NIH)/National Institute of Mental Health (R34MH113539) and NIH/National Center for Research Resources Colorado Clinical and Translational Science Institute (UL1RR025780), but the sponsor had no role in review or approval of the manuscript for publication. Its contents are the sole responsibility of the authors and do not necessarily represent official NIH views. The authors thank Silvia Raghunath for her assistance with participant enrollment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representative screenshots of Lock to Live.
[PDF File (Adobe PDF File), 343 KB - jmir_v22i1e16253_app1.pdf]

Multimedia Appendix 2
Firearm storage options chosen by intervention participants after viewing Lock to Live (Multiple responses are allowed. Both family, friend, or neighbor and lock box were selected as defaults. Participants could unselect them if they chose).
[PNG File , 22 KB - jmir_v22i1e16253_app2.png]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 97 KB - jmir_v22i1e16253_app3.pdf]

References

1. The Joint Commission. Detecting and treating suicide ideation in all settings. Sentinel Event Alert 2016 Feb 24(56):1-7. [Medline: 26915165]
2. Mann JJ, Michel CA. Prevention of firearm suicide in the United States: what works and what is possible. Am J Psychiatry 2016 Oct 1;173(10):969-979. [doi: 10.1176/appi.ajp.2016.16010069] [Medline: 27444796]
3. Office of the Surgeon General (US), National Action Alliance for Suicide Prevention (US). 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action. A Report of the US Surgeon General and of the National Action Alliance for Suicide Prevention. Washington, DC: US Department of Health & Human Services (US); 2012.
4. Spicer RS, Miller TR. Suicide acts in 8 states: incidence and case fatality rates by demographics and method. Am J Public Health 2000 Dec;90(12):1885-1891. [doi: 10.2105/ajph.90.12.1885] [Medline: 11111261]

5. Miller M, Azrael D, Hemenway D. The epidemiology of case fatality rates for suicide in the northeast. Ann Emerg Med 2004 Jun;43(6):723-730. [doi: 10.1016/S0196064404000691] [Medline: 15159703]
6. Weinberger SE, Hoyt DB, Lawrence HC, Levin S, Henley DE, Alden ER, et al. Firearm-related injury and death in the United States: a call to action from 8 health professional organizations and the American Bar Association. Ann Intern Med 2015 Apr 7;162(7):513-516. [doi: 10.7326/M15-0337] [Medline: 25706470]
7. Bulger EM, Kuhls DA, Campbell BT, Bonne S, Cunningham RM, Betz M, et al. Proceedings from the medical summit on firearm injury prevention: a public health approach to reduce death and disability in the US. J Am Coll Surg 2019 Oct;229(4):415-30.e12. [doi: 10.1016/j.jamcollsurg.2019.05.018] [Medline: 31108194]
8. American Foundation for Suicide Prevention. 2016 Aug 17. Nation's Largest Suicide Prevention Organization Launches Suicide Prevention and Firearm Pilot Program URL: https://afsp.org/nations-largest-suicide-prevention-organization-launches-suicide-prevention-firearm-pilot-program/ [accessed 2019-12-26]
9. National Shooting Sports Foundation.: National Shooting Sports Foundation; 2016 Aug 22. NSSF-AFSP Suicide Prevention Partnership URL: http://www.nssfblog.com/nssf-afsp-suicide-prevention-partnership/ [accessed 2019-12-26]
10. Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS) URL: http://www.cdc.gov/injury/wisqars/index.html [accessed 2020-01-01]
11. Knoepke CE, Allen A, Ranney ML, Wintemute GJ, Matlock DD, Betz ME. Loaded questions: internet commenters' opinions on physician-patient firearm safety conversations. West J Emerg Med 2017 Aug;18(5):903-912 [FREE Full text] [doi: 10.5811/westjem.2017.6.34849] [Medline: 28874943]
12. Wintemute GJ, Betz ME, Ranney ML. Yes, you can: Physicians, patients, and firearms. Ann Intern Med 2016 Aug 2;165(3):205-213. [doi: 10.7326/M15-2905] [Medline: 27183181]
13. Betz ME, Kautzman M, Segal DL, Miller I, Camargo CA, Boudreaux ED, et al. Frequency of lethal means assessment among emergency department patients with a positive suicide risk screen. Psychiatry Res 2018 Feb;260:30-35 [FREE Full text] [doi: 10.1016/j.psychres.2017.11.038] [Medline: 29169036]
14. Betz ME, Miller M, Barber C, Beaty B, Miller I, Camargo CA, et al. Lethal means access and assessment among suicidal emergency department patients. Depress Anxiety 2016 Jun;33(6):502-511 [FREE Full text] [doi: 10.1002/da.22486] [Medline: 26989850]
15. Petrik ML, Gutierrez PM, Berlin JS, Saunders SM. Barriers and facilitators of suicide risk assessment in emergency departments: a qualitative study of provider perspectives. Gen Hosp Psychiatry 2015;37(6):581-586. [doi: 10.1016/j.genhosppsych.2015.06.018] [Medline: 26208868]
16. Walters H, Kulkarni M, Forman J, Roeder K, Travis J, Valenstein M. Feasibility and acceptability of interventions to delay gun access in VA mental health settings. Gen Hosp Psychiatry 2012;34(6):692-698. [doi: 10.1016/j.genhosppsych.2012.07.012] [Medline: 22959420]
17. Betz ME, Brooks-Russell A, Brandspigel S, Novins DK, Tung GJ, Runyan C. Counseling suicidal patients about access to lethal means: Attitudes of emergency nurse leaders. J Emerg Nurs 2018 Sep;44(5):499-504 [FREE Full text] [doi: 10.1016/j.jen.2018.03.012] [Medline: 29704978]
18. Betz ME, Knoepke CE, Siry B, Clement A, Azrael D, Matlock DD. 2018. Lock to Live URL: http://lock2live.org
19. Betz ME, Knoepke CE, Siry B, Clement A, Azrael D, Ernestus S, et al. 'Lock to Live': development of a firearm storage decision aid to enhance lethal means counselling and prevent suicide. Inj Prev 2019 Sep;25(Suppl 1):i18-i24. [doi: 10.1136/injuryprev-2018-042944] [Medline: 30317220]
20. O'Connor AM, Cranney A. Patient Decision Aids - Ottawa Hospital Research Institute. 1996. User Manual - Acceptability URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf [accessed 2019-12-26]
21. O'Connor AM. Patient Decision Aids - Ottawa Hospital Research Institute. 2006. Ottawa Decision Support Framework to Address Decisional Conflict URL: https://decisionaid.ohri.ca/docs/develop/ODSF.pdf [accessed 2020-01-01]
22. O'Connor AM. Patient Decision Aids - Ottawa Hospital Research Institute. 1993. User Manual - Decisional Conflict Scale URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf [accessed 2020-01-01]
23. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
24. Levy AG, Scherer AM, Zikmund-Fisher BJ, Larkin K, Barnes GD, Fagerlin A. Assessment of patient nondisclosures to clinicians of experiencing imminent threats. JAMA Netw Open 2019 Aug 2;2(8):e199277 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9277] [Medline: 31411716]
25. Karras E, Stokes CM, Warfield SC, Barth SK, Bossarte RM. A randomized controlled trial of public messaging to promote safe firearm storage among US military veterans. Soc Sci Med 2019 Nov;241:112205. [doi: 10.1016/j.socscimed.2019.03.001] [Medline: 31387766]
26. Anestis MD, Butterworth SE, Houtsma C. Perceptions of firearms and suicide: The role of misinformation in storage practices and openness to means safety measures. J Affect Disord 2018 Feb;227:530-535. [doi: 10.1016/j.jad.2017.11.057] [Medline: 29169121]

27. Anestis MD, Houtsma C. What I have is what I am: Differences in demographics, suicidal thoughts and behaviors, and firearm behavior and beliefs between firearm owners who do and do not primarily identify as firearm owners. J Psychiatr Res 2019 Sep;116:157-165. [doi: 10.1016/j.jpsychires.2019.06.017] [Medline: 31247359]
28. Roszko PJ, Ameli J, Carter PM, Cunningham RM, Ranney ML. Clinician attitudes, screening practices, and interventions to reduce firearm-related injury. Epidemiol Rev 2016;38(1):87-110. [doi: 10.1093/epirev/mxv005] [Medline: 26905894]
29. Kelly T, Brandspigel S, Polzer E, Betz ME. Firearm Storage Maps: A Pragmatic Approach to Reduce Firearm Suicide During Times of Risk. Ann Intern Med 2020 Jan 21. [doi: 10.7326/M19-2944] [Medline: 31958811]
30. Siry B, Betz M, Saunders T, Knoepke C, Simpson S, Matlock D, et al. A pilot trial of suicidal adults with in emergency departmentsnrollment challenges and strategies related to firearm access. SAGE Research Methods Cases 2020 (in press).

Abbreviations

DCS: Decisional Conflict Scale
ED: emergency department
L2L: Lock to Live
LMC: lethal means counseling
NIH: National Institutes of Health
RA: research assistant
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by R Ho, M Ranney; comments to author 12.11.19; revised version received 15.11.19; accepted 15.12.19; published 29.01.20.

Please cite as:
Betz ME, Knoepke CE, Simpson S, Siry BJ, Clement A, Saunders T, Johnson R, Azrael D, Boudreaux ED, Omeragic F, Adams LM, Almond S, Juarez-Colunga E, Matlock DD
An Interactive Web-Based Lethal Means Safety Decision Aid for Suicidal Adults (Lock to Live): Pilot Randomized Controlled Trial
J Med Internet Res 2020;22(1):e16253
URL: http://www.jmir.org/2020/1/e16253/
doi:10.2196/16253
PMID:32012056

Original Paper

A User-Centered Approach to an Evidence-Based Electronic Health Pain Management Intervention for People With Chronic Pain: Design and Development of EPIO

Ingrid Konstanse Ledel Solem[1,2], MSc; Cecilie Varsi[1], PhD; Hilde Eide[1,3], PhD; Olöf Birna Kristjansdottir[1,4], PhD; Elin Børøsund[1], PhD; Karlein M G Schreurs[5], PhD; Lori B Waxenberg[6], PhD; Karen E Weiss[7], PhD; Eleshia J Morrison[8], PhD; Mette Haaland-Øverby[1,4], BSc; Katherine Bevan[9], BSc; Heidi Andersen Zangi[10,11], PhD; Audun Stubhaug[2,12], PhD; Lise Solberg Nes[1,2,8], PhD

[1]Department of Digital Health 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, Gainesville, FL, United States
[7]Department of Anesthesiology and Pain Medicine, School of Medicine, University of Washington, Seattle, WA, United States
[8]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States
[9]Center for Learning and Mastery, Bærum Hospital, Vestre Viken Hospital Trust, Bærum, Norway
[10]National Advisory Unit on Rehabilitation in Rheumatology, Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway
[11]Faculty of Health, VID Specialized University, Oslo, Norway
[12]Regional Advisory Unit on Pain, Oslo University Hospital, Oslo, Norway

Corresponding Author:
Lise Solberg Nes, PhD
Department of Digital Health Research
Division of Medicine
Oslo University Hospital
Pb 4950 Nydalen
Oslo, N-0424
Norway
Phone: 47 91332341
Email: lise.solberg.nes@rr-research.no

Abstract

Background: Chronic pain conditions are complicated and challenging to live with. Electronic health (eHealth) interventions show promise in helping people cope with chronic illness, including pain. The success of these interventions depends not only on the technology and intervention content but also on the users' acceptance and adherence. Involving all stakeholders (eg, patients, spouses, health care providers, designers, software developers, and researchers) and exploring their input and preferences in the design and development process is an important step toward developing meaningful interventions and possibly strengthening treatment outcomes.

Objective: The aim of this study was to design and develop a user-centered, evidence-based eHealth self-management intervention for people with chronic pain.

Methods: The study employed a multidisciplinary and user-centered design approach. Overall, 20 stakeholders from the project team (ie, 7 researchers, 5 editors, 7 software developers, and 1 user representative), together with 33 external stakeholders (ie, 12 health care providers, 1 health care manger, 1 eHealth research psychologist, and 17 patients with chronic pain and 2 of their spouses) participated in a user-centered development process that included workshops, intervention content development, and usability testing. Intervention content was developed and finalized based on existing evidence, stakeholder input, and user testing. Stakeholder input was examined through qualitative analyses with rapid and in-depth analysis approaches.

Results: Analyses from stakeholder input identified themes including a need for reliable, trustworthy, and evidence-based content, personalization, options for feedback, behavioral tracking, and self-assessment/registration as factors to include in the intervention. Evidence-based intervention content development resulted in one face-to-face introduction session and 9 app-based educational and exercise-based modules. Usability testing provided further insight into how to optimize the design of the intervention to the user group, identifying accessibility and a simple design to be essential.

Conclusions: The design and development process of eHealth interventions should strive to combine well-known evidence-based concepts with stakeholder input. This study, designing and developing the pain management intervention EPIO, illustrates how a stakeholder-centered design approach can provide essential input in the development of an eHealth self-management intervention for people with chronic pain.

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

(J Med Internet Res 2020;22(1):e15889) doi:10.2196/15889

KEYWORDS

Web-based interventions; eHealth; mobile apps; evidence-based, user-centered design approach; service design; chronic pain; cognitive behavioral therapy; acceptance and commitment therapy

Introduction

Background

Chronic pain conditions are often multifaceted and difficult to manage, involving physiological and psychological and social challenges for those affected [1-4]. Given this complexity, chronic pain can also be perceived as unavoidable, unmanageable, and challenging to disengage from [5].

Psychoeducational individual or group-based interventions with cognitive behavioral (ie, cognitive behavioral therapy; CBT) [6] and/or acceptance and commitment (ie, acceptance and commitment therapy; ACT [7]) approaches supporting self-management have been shown to be effective. CBT entails an integrative approach combining cognitive and behavioral change techniques, focusing on challenging and changing unhelpful thoughts and behaviors with a goal-oriented, problem-solving approach and enhancement of coping strategies [8]. ACT was initially proposed as a new generation of CBT, focusing on the role of acceptance and mindfulness rather than cognitive change; aiming to increase psychological flexibility; and centering around acceptance, awareness of the moment, and a commitment to values and direction [9]. Both types of interventions can improve a person's quality of life, pain acceptance, functioning, and self-efficacy while also having the potential to reduce pain, depressive symptoms, and distress [10-18]. Unfortunately, such in-person interventions are not always offered or available [5], and multiple barriers to attendance may be present for people living with chronic pain conditions.

Electronic health (eHealth) interventions, referring to health-related interventions distributed through technology, have shown promising results in helping people cope with health-related issues and chronic illness, including pain [19-25]. eHealth interventions have the potential to offer patients easier access to illness management when convenient and most needed and at the patients' own pace [26,27]. Given the potential for flexibility in use, eHealth interventions may introduce more cost-effective treatment options, supplementing usual care and even reducing the need for direct involvement from health care providers [28,29]. This could be particularly important when dealing with chronic illness, such as pain, as individuals living with chronic illness usually have the need for, and responsibility of, day-to-day management of their own illness [13]. Several studies have pointed to great potential for the use of eHealth in chronic pain management [21,23,24,27,30,31]. However, a number of challenges have been associated with such interventions. One significant challenge is the lack of guidance and involvement of health care providers and intended users in the development process [21,27,32-37]. This has resulted in a gap between the commercial and scientific sides of eHealth, with products often being developed in response to technological innovations rather than evidence-based knowledge and/or user needs [27,33,38]. There is a need for more attention on how to develop and translate or transform existing face-to-face interventions into electronic formats at the same time focusing on the actual needs of patients with pain [27]. Researchers have made recommendations for utilization of a more user-centered design approach, ideally involving all stakeholders (eg, patients, health care providers and caregivers, pain and eHealth researchers, and designers and information technology [IT] developers) in the eHealth intervention development process from the early idea initiation to the final intervention evaluation [5,27,39]. Despite these recommendations, end users and other stakeholders are still rarely involved in the early development process of eHealth interventions [32,39]. This could potentially be at least partially because of a lack of information and frameworks on *how* to involve stakeholders in the development process [39].

Additional challenges include low adherence and high attrition rates in eHealth interventions [39-43]. For eHealth interventions and development processes to be successful, a focus on the *entire* person (ie, a holistic view), including relationships, context, and intervention setting, is necessary [39,44,45]. This includes identifying facilitators and barriers for use and exploring implementation issues from an early stage on [46-49]. Addressing these issues will likely contribute to development of more user friendly, meaningful interventions for patients and can potentially improve treatment outcomes for patients living with chronic pain [27,39,44,46].

This study is part of a larger project aiming to design, develop, and test the effectiveness of a user-centered, evidence-based

eHealth self-management intervention for adults with chronic pain (clinical trial registration: NCT03705104). In the first step of the larger project, users' everyday challenges and attitudes toward eHealth technology, as well as their needs and requirements for a potential eHealth pain management intervention, were explored through individual interviews with people with chronic pain and their spouses [50]. Individual interviews with health care providers have also been conducted focusing on the same issues (to be published elsewhere). Patients and spouses in the initial study reported a broad spectrum of everyday challenges, including physical, psychological, and social challenges, such as fatigue, grief, guilt, and social- and work isolation, and participants anticipated that eHealth technology would be a positive and accessible option for pain management support [50]. The study found that patients' needs in relation to an eHealth pain management intervention can be summarized in 3 main areas: (1) need for reliable knowledge about pain and pain management, including access to useful coping skills and exercises; (2) support in finding balance in everyday life, physically and mentally, through increased awareness and simple documentation (ie, ability to track variables such as sleep, mood, physical activity, and pain); and (3) social support, including peer support forums and advice on how to communicate with others, such as family, friends, and health care providers [50].

Objectives

Building on the recent findings [50], this study aimed to design and develop a user-centered, evidence-based eHealth self-management intervention for people with chronic pain. This paper includes descriptions and results from the development process, including results from workshops with all involved stakeholders (eg, patients, spouses, health care providers, researchers, software developers, and user representatives), intervention content development and usability testing. The ultimate goal was to develop an evidence-based intervention that was acceptable to users (ie, well received, suitable, user friendly, attractive, and meeting needs) [51] and had the potential to produce changes in quality of life for people with chronic pain.

Methods

Study Design

The design and development process entailed a multidisciplinary and user-centered design approach [39,52,53]. The project utilized well-established cognitive behavioral pain management concepts shown to be effective for people with chronic pain [11,14,16,54,55] and incorporated concepts of a participatory design approach [56] to ensure that the intervention would be acceptable (ie, well received, suitable, user friendly, and attractive) and designed in line with patients' needs and context of use.

The intervention development was led by the study principal investigator (PI; LSN), who is a clinical psychologist with health psychology specialization and long-standing experience within chronic pain and cognitive behavioral treatment approaches for medical patients. The multidisciplinary project team consisted of health care researchers (ie, PhD-level psychologists and registered nurses), an editorial group ensuring that content and material was presented in an understandable way, a software team (ie, software developers and a designer), and a user representative (ie, person with chronic pain experience). The team met weekly (sometimes more often) during the development process. See Table 1 for an overview of the project team members with their project-related expertise.

Table 1. Overview of multidisciplinary project team (N=20).

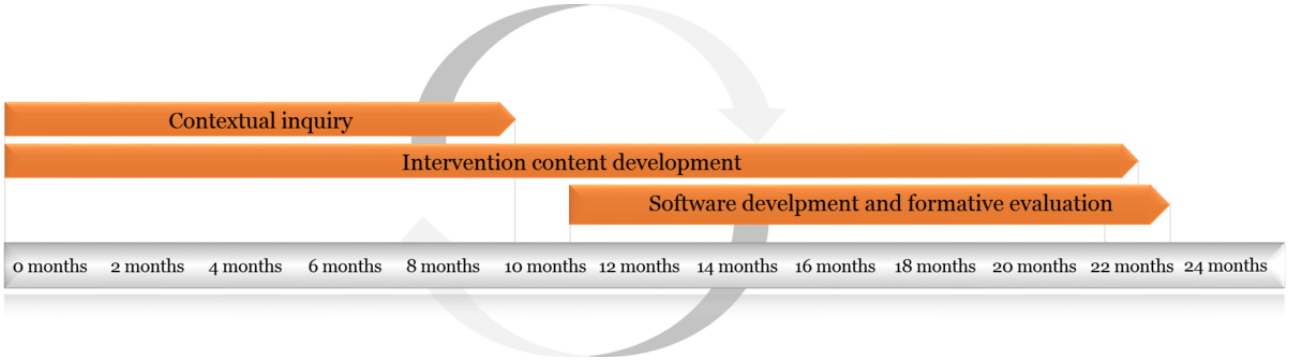
Grouping	Total number, n	Pain expertise, n	Electronic health expertise, n	Licensed health care providers, n
Health care researchers	7	6	5	5
Editorial group	5	0	5	2
Software team	7	0	7	0
User representative	1	1	0	0

Patients and other stakeholders, including spouses of patients with chronic pain, and collaborative partners (ie, interdisciplinary health care providers, health care managers, and researchers working at collaborating institutions such as hospitals, municipality health care services, and universities), all with long-standing experience on chronic pain issues, were also involved in the development process. Service design methods, utilizing a user-centered, sequenced, cocreative, and holistic development approach [57] were used to facilitate co-design and high user engagement throughout the development process.

The pain management intervention was developed in an iterative process involving systematic evaluation throughout as suggested by the Center for eHealth Research and Disease Management comprehensive roadmap approach for the uptake and impact of eHealth technologies [39].

The intervention was developed in iterative processes, as shown in Figure 1, through a combination of (1) contextual inquiry and co-design processes collecting input from people with chronic pain, their spouses, health care providers, and eHealth experts; (2) intervention content development, where content was developed by members of the project team based on evidence-based CBT and ACT concepts for chronic pain self-management; and (3) iterative software development and formative evaluation, including low- and high-fidelity prototypes development and usability testing.

Figure 1. Study overview and development timeline.



Recruitment

To be eligible for study participation, the *patients* had to be 18 years or older, having experienced chronic pain for 3 months or more, and had to be able to communicate in Norwegian. Patients were encouraged to participate regardless of age (≥18 years) and gender. *Spouses* had to be married or cohabitating with one of the participating patients.

Patients and spouses were recruited through study information published on the Web as well as through national project collaborators and local and national pain associations. In addition, patients and spouses participating in the initial interviews investigating patients' needs and requirements for eHealth interventions [50] were invited to participate in this study, as they already had experience with the topic and could potentially add another layer to the design discussions.

Data Collection and Analysis

Data were collected from pain management courses, workshops, and usability testing. Then, to ensure that the collected material provided essential input into the ongoing development process, collected data were first analyzed by means of rapid analysis [58]. This included summarizing data from voice recordings and recorded notes before sharing and discussing the material in the project group (including the development team). Following this process, to ensure a thorough scientific development and that no themes were overlooked, a more in-depth analysis of the material took place using directed content analysis [59]. In this process, the material was coded into predefined categories, containing development suggestions and requirements from the participants, including input for content, design, and functionality.

Intervention Development

Contextual Inquiry: Data Collection

A contextual inquiry [39] initiated the development process to gather information about the intended users, their needs and requirements for acceptability, and the environment in which the intervention was intended to fit, building on previously gathered information [50].

Pain Management Course Observations

To gain an insight into health care services offered in the study area, as well as to gain additional information related to patients' needs and experiences, the first (ILS) and second (CV) authors observed 5 different pain management courses available in local patient education centers and hospitals. Notes and summaries were recorded during and after the course observations. The information gathered was summarized into 3 categories: (1) topics covered in the courses, (2) information shared by course participants (ie patients), and (3) themes that potentially could be incorporated into the app. This information was shared with the project team to increase knowledge about the patient group and the pain management courses with the rest of the project team members.

Developing Personas and Patient Journey Maps

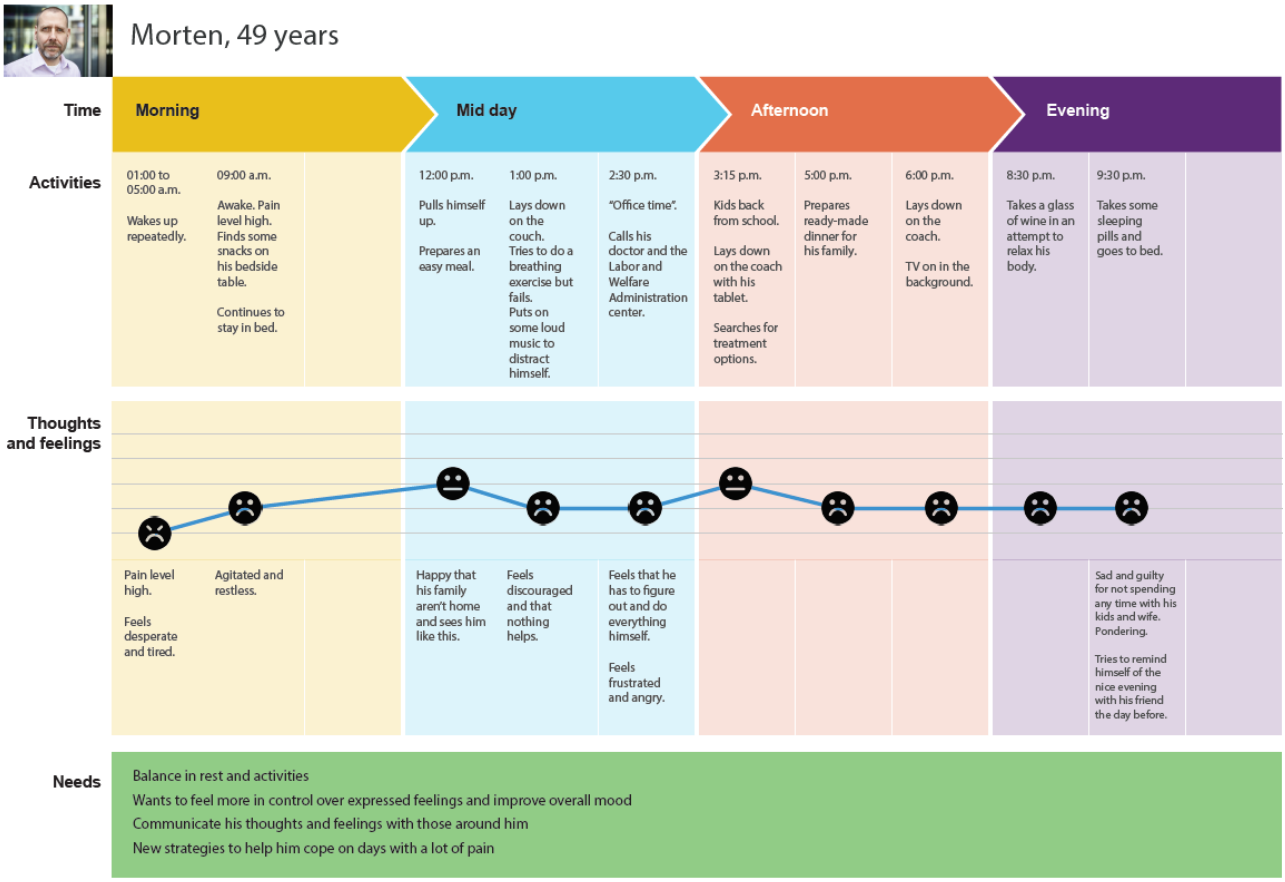
On the basis of existing research recommendations [39,60], service design methods [57] were used to facilitate user engagement. Five personas (ie, fictional but representative patient profiles; see example in Figure 2) and 2 *patient journey maps* (ie, *roadmaps* inspired by customer journey maps [57]; see Figure 3) visualizing typical days in the patients' lives were developed based on previous findings [50] for the design and development process. Personas included background information (ie, stories to give each persona more depth), coping skills and everyday challenges, an overview of technology skills, and the persona's needs and requirements in relation to the eHealth intervention. The journey maps described a *typical day* in a patient's life. Personas and journey maps informed the development process and project team members about *typical* end users and their daily challenges, needs, and requirements to bring the *patients' voices* to the forefront in the development process. Personas and journey maps were also used in the upcoming workshops as illustrative scenarios that the participants could use when discussing possible design and functionality options. For illustrations, see Figures 2 and 3.

Workshops

Contributing to the contextual inquiry, stakeholders were invited to discuss their needs, requirements for acceptability, ideas, and possible challenges related to the eHealth project. This was done successively through (1) 5 workshops with participants from collaborating institutions and the project team and (2) 3 stakeholder workshops involving all stakeholders (ie, people with chronic pain, spouses, health care providers, researchers, editors, and the software team). Service design methods, including the use of scenario tasks, personas, and journey maps [57], were used to facilitate participant engagement. The workshops are described in the following sections.

Figure 2. Illustration of a study persona.

Figure 3. Illustration of a patient journey map during a typical day with pain.



Workshops With Collaborating Partners

First, health care providers, a health care manager, and an eHealth research psychologist from collaborating institutions (ie, hospitals, universities, and primary health care services) were invited to participate in workshops together with the project team.

In the first workshop, participants were separated into smaller groups consisting of health care providers with a variety of professional backgrounds (eg, registered nurses, psychologists, and social workers) and individuals from the project team (including eHealth researchers and software developers). Group discussions were timed and led by group facilitators from the project team. Notes from the group discussions were recorded and materials such as post-it notes and drawings from group tasks were collected. The material was categorized by the first author (ILS) and another project team member using a rapid analysis approach [58] to ensure rapid and continuous input on the development process, broadly sorting the material into idea clusters based on content and similarities.

Information gathered in the first workshop provided guidance for the subsequent 4 workshops, which focused more specifically on the development process (ie, content and software development), with workshop 2 focusing on adherence and design elements and workshops 3 to 5 focusing on content development and how to best present the psychoeducational content. Notes were taken and analyzed using a rapid analysis approach [58].

Workshops With All Stakeholders

Building on workshops with collaborative partners, all stakeholders (ie, people with chronic pain, spouses, health care providers, researchers, editors, and the software team) participated in 1 out of 3 stakeholder workshops. The main purpose of these workshops was to elicit ideas on design and content features and to further explore users' requirements for an eHealth pain management intervention, with each workshop informing the next. The first stakeholder workshop was arranged with only patients participating, together with members from the project team (ie, the designer and 2 researchers, including the first author). In stakeholder workshops 2 and 3, spouses and health care providers from collaborating institutions were also included in addition to editors and software team members. Participants were divided into multidisciplinary groups of 5 to 6. A brief presentation of the personas and journey maps developed initiated each workshop to provide all stakeholders with an overview of findings in the development process so far and to provide a collective understanding of the target patient group. The personas were also updated during the workshops based on participant feedback.

In the first workshop, more time was spent on discussing the personas, whereas stakeholder workshops 2 and 3 focused more on design and functionality aspects. As a starting exercise, participants were asked to reflect upon what makes an app *good or bad* and to discuss usability and acceptability aspects within their groups. Potential design features were then briefly presented to support stakeholders when participating in the design discussions. The participants were then asked to reflect and discuss which design features and elements were most important to them in a priority task, where participants had to choose between different design elements and features. The final part of the workshops included a collective design task, using scenarios, the personas, and patient journey maps to discuss possibilities related to an eHealth self-management intervention (ie, content, functionality, design, and usage) before finally looking at potential barriers for use.

All stakeholder workshops were facilitated by the first author in collaboration with other project team members (KH, HS, MW, JM, and YI), and each group discussion was audio taped. The material was first summarized by the first author using rapid analysis [58] and focusing on ensuring that the material provided essential input into the ongoing development process. The material was later transcribed verbatim to conduct a more thorough analysis using directed content analysis [59] to ensure a thorough scientific process in material identifications. The material was first sorted into broad categories representing requirements for (1) content, (2) functionality, (3) design, and (4) barriers for use. Data were extracted and compared across the different workshops, looking for similarities and differences in the material.

Intervention Content Development

A vital goal in this study was to identify evidence-based topics and aspects from recognized cognitive behavioral and acceptance and commitment pain management strategies; then, develop the intervention content and integrate findings and content with a user-centered approach, and subsequently, modify findings to create a new technology-based pain management intervention for people living with chronic pain.

Intervention content development was led by the project PI (LSN), in close collaboration with the other experienced pain management project team members (ie, co-authors KS, LW, EM, KW, HE, and OK), assisted by 3 editors (MW, EB, and HS) and the project-specific user representative. Following the initial workshop with collaborating partners, the intervention content development group first examined the existing literature in the clinical and research area, then discussed the findings and compared notes also based on clinical pain management experience within the group. Intervention content material was then developed based on evidence-based topics and aspects from recognized CBT and ACT pain management [61-64], tested and user tested, and then adjusted and adapted accordingly in continuous iterations. Throughout the process, intervention content was shaped for an electronic format to support usability and ease of use for the end users. The intervention content development underwent numerous iterations (ie, number of iterations varied depending on topic/module) to certify that it used appropriate, therapeutic language; was presented in brief and easily understandable sentences; and was suitable for small screens.

Software Development and Formative Evaluation

On the basis of content development and stakeholder input, a low-fidelity paper prototype of the software was developed. The prototype was tested within the development team with involvement from eHealth experts and the project team, then

adjusted and implemented electronically to simulate the app idea. To strengthen acceptability, the simulated prototype was subsequently tested by the project team user representative, hospital-employed healthy volunteers, and 1 external patient before full-scale usability testing.

Technical Architecture

EPIO is distributed as a native app for iOS and Android through the official app stores, and it is implemented using Web technology in a Cordova container. All information stored locally is encrypted with the Advanced Encryption Standard (AES) algorithm in Galois/counter mode before it is written to a local SQLite (a relational database management system) instance. The key used for encryption is 256 bits long, and it is generated the first time the app runs. Between invocations of the app, it is wrapped using the AES- key-wrapped algorithm with a wrapping key derived from the user's personal identification number (PIN) and stored on the device's keychain. As the keychain itself requires the device to be protected with a PIN, the role of the app's own PIN is to enable the user to secure the app even when using it on a shared device. Usage logs (navigation, use, and use of functionality) and self-assessments are sent over an encrypted channel to a secure server for later analysis by the research project staff.

Technical decisions were executed only after discussions in the project team (ie, researchers, health care personnel, eHealth experts, software design, and developers and user representatives).

Usability Testing

Building on feedback and discussions within the project team, high-fidelity prototypes were developed, including a start page, menu page, and intervention modules. A diverse group of users (ie, variety in age/gender), including hospital-employed healthy volunteers and people with chronic pain, participated in the testing.

The high-fidelity prototype usability tests were videotaped and conducted face to face by a facilitator (ie, editor/eHealth expert) and an observant (ie, either the first author, the designer, or another project team member). A think aloud methodology [57] was used to actively engage the participants and elicit continuous feedback, with participants describing their actions and immediate thoughts for each step. The observer took notes throughout the testing. Summaries of observations were completed and transferred into a table by the facilitator and observer following each testing, containing information related to (1) usability issues, (2) possible solutions, (3) who reported this issue (ie, number of users), and (4) other input. This provided a rapid and continuous yet structured feedback into the development process. As a supplement toward establishing acceptability, participants completed Sauro's System Usability Scale (SUS) [65]. This was done at the end of the usability testing after the facilitator and the observer had left the room. The SUS measures usability and satisfaction on a scale from 1 (strongly disagree) to 5 (strongly agree).

The summaries from the usability testing were discussed within the project team and new sketches and decisions for the next development phase were conducted, with the project team discussing and prioritizing changes. The collected material was later examined more in depth through content analysis [59] to potentially identify themes overlooked in the initial rapid analysis [58]. In this process, the material was sorted into broad categories looking at (1) usability and flow, (2) functionality and customization, (3) intervention content, and (4) design and language.

Security and Privacy Considerations

The intervention program was developed at a major medical center in Northern Europe. The design and development were in accordance with the European General Data Protection Regulations of 2018. The study, including a risk assessment analysis of the app, was approved by the institution's Department of Information Safety and the institutional review board (approval number: 2017/6697). Informed consent will be obtained from all users of the app-based program.

Results

Participants

A total of 33 participants participated in the study design and development process (ie, workshops and content and software development), including 12 health care providers, 1 health care manager, and 1 eHealth research psychologist from collaborating institutions, as well as 17 patients and 2 of their spouses together with the project team (see Table 1 in the Methods section). For details, please see Table 2 for collaborating partners' background and expertise, Table 3 for patient demographics, and Figure 4 for a complete overview of the intervention development process, including activities and participation.

All participants, 3 men and 11 women, from collaborating institutions had extensive experience working with people living with chronic pain. They represented a variety of professional backgrounds, with the majority working as licensed psychologists within chronic pain management. In addition, 9 had experience in research and 3 had eHealth expertise.

The majority of the patients had experience with a variety of treatments, ranging from primary care and physical therapy to more specialized treatments and rehabilitation in secondary and tertiary care settings. All participating patients owned a smartphone and had access to a computer and/or tablet. Most of the patients used apps daily, which were either installed by themselves or someone in their family, though several were not familiar with the concept of apps and did not know the difference between a webpage and an app. Many of the patients participated first in 1 out of the 3 workshops and then also later during software development and formative evaluation. The 2 male spouses had been part of the initial interviews [50].

Table 2. Overview of background and experience of participants from collaborating institutions (N=14).

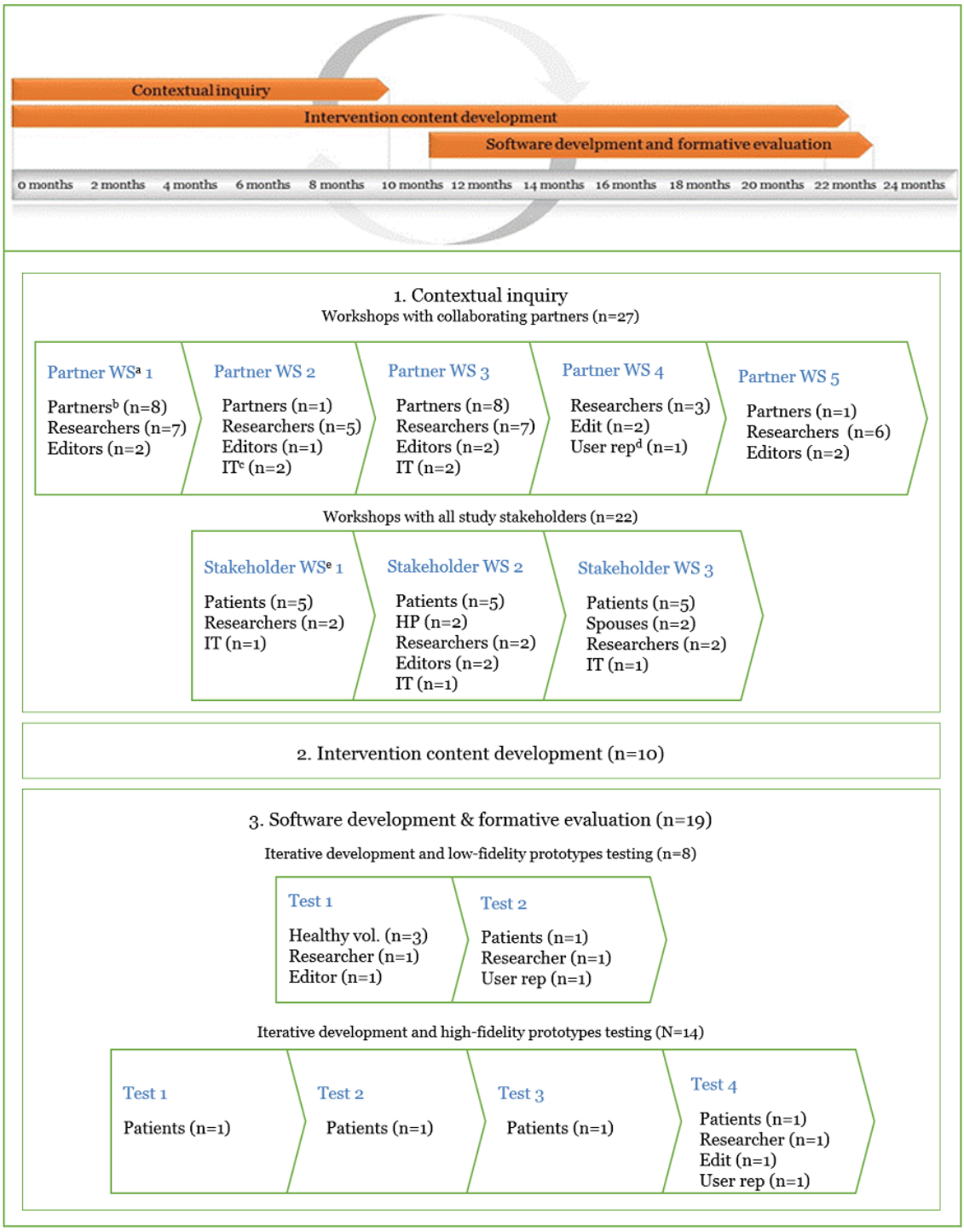
Health care background	Total number, N	Pain expertise, n	eHealth expertise, n	Research expertise, n
Nurse	1	1	0	1
Psychologist	7	7	1	5
Physician	1	1	0	1
Social worker	2	2	1	1
Occupational therapist	1	1	0	0
Nonlicensed partner	2[a]	1	1	1

[a]Health care manager and eHealth research psychologist.

Table 3. Patient demographics (N=17).

Characteristics	Values, n (%)
Gender	
Male	2 (12)
Female	15 (88)
Age (years), mean age=51 years	
20-35	2 (12)
36-50	5 (29)
51-59	7 (41)
60-75	3 (18)
Type of pain (primary diagnosis reported)	
Neck and/or back pain	5 (29)
Nerve pain/neuropathic pain	5 (29)
Fibromyalgia	3 (18)
Migraine	2 (12)
Others	2 (12)
Reported years living with pain	
0-5	2 (12)
6-10	6 (35)
11-17	3 (18)
18-26	4 (24)
≥27	2 (12)
Employment status	
Working/studying full time	4 (24)
Working/studying part time	3 (18)
On disability benefits	8 (47)
Retired	1 (6)
Nonworking	1 (6)

Figure 4. Overview of study timeline, intervention development process, activities, and participation. Partner WS: collaborating partners; Partners: collaborating partners; IT: person(s) from software team (ie, developers and designer); User rep: user representative; Stakeholder WS: stakeholder workshop. The test facilitator and observers are not counted as participants and included in the n for software development and formative evaluation.



Pain Management Course Observations

The 5 pain management courses observed covered topics such as pain physiology, coping skills, and psychosocial challenges. They had different structures, the majority being 1- or 2-day courses and 1 occurring weekly over a period of 3 months. The pain management courses covered topics such as pain physiology, coping skills, and psychosocial challenges. The level of information shared by patients during the courses varied, depending on whether the type of course they attended had allocated time for discussions or mainly focused on educational information. However, the patients did describe a general need for more peer support in their everyday lives. Topics otherwise emphasized by participating patients included grief, guilt, anxiety, and negative thoughts as well as issues related to communication with health care providers, spouses/family, and others. Participants generally seemed pleased with the courses they attended, even though several patients emphasized the need for more continuous support and help in everyday life.

Workshops With Collaborating Partners

Workshop findings showed that participants from collaborating institutions (ie, health care providers and eHealth experts, n=14) were generally positive toward an eHealth intervention for people with chronic pain and saw it as an opportunity to make psychological and educational treatment more accessible to patients during and after treatment. Health care providers pointed to the importance of reliable, evidence-based educational information. Please see Table 4 for content needs and rationales expressed during these workshops.

Table 4. Results from workshops with collaborating partners (N=14).

Content needs and topics	Rationale
Reliable, trustworthy, evidence-based knowledge	Provide evidence-based, trustworthy information to patients, giving them a better alternative to online forums and other nonscientific channels
Focus on psychological health	Increase patient's awareness of the association between chronic pain and psychological challenges
Activity pacing	Support patients in implementing activity pacing strategies in everyday life, including through exercises
Self-assessment/registrations	Increase awareness about personal activities and positive/negative health behavior (eg, amount of sleep and physical activity)
Communication	Include advice on how to best communicate personal struggles, potentially adding direct contact with health care professionals as a functionality
Social support	Adding some form of option for social contact with peers

Potential barriers in the development process, as well as for the final product, were identified as (1) time challenges, referring to the amount of time it takes to develop eHealth interventions and whether or not health care providers had time to participate in the development process; (2) defining the most relevant and beneficial intervention content; (3) patient involvement in the development process; and (4) privacy and security issues.

Workshops With All Study Stakeholders

The 3 stakeholder workshops identified that patients (n=17), spouses (n=2), and health care providers (n=2) supported many of the same thoughts and ideas as health care professionals and other collaborating partners participating in the initial 5 workshops. The patients particularly emphasized the need for an intervention that gave them positive input in their daily lives. They did not want reminders of what they could *not* do, for example, being asked to do *impossible* exercises or being asked to set *unrealistic* personal goals. Feeling guilty, grief, achieving balance in everyday life, getting support, being present and being useful while taking care of oneself were topics mentioned by all participating groups. Content topics suggested and discussed by participants included updates on recent scientific findings; information about health-promoting behavior (ie, sleep, nutrition, and physical activity); and advice on how to prioritize and set limits, support and information on acceptance, and exercises promoting energy and awareness, such as breathing and relaxation exercises.

Regarding the end product/solution and what it should look like, the participants had a broad range of functionality suggestions and demands. All external stakeholders pointed to personalization, that is, adjusting the intervention based on individual needs and personal preferences. For example, customization and simple behavior trackers were pointed out by some as important features for a chronic pain eHealth intervention. At the same time, several patients, spouses, and health care providers emphasized that too many options for choice could potentially be perceived as overwhelming. The use of gamification elements (ie, application of game-playing elements such as avatars, points, and badges) was viewed as a potentially important and motivating option by collaborating partners and the software team. However, the participating patients did not identify gamification as important compared with other potential elements and features of the solution. Too many or too bright colors, cartoons, or sound effects were described by several patients as potentially challenging for them, especially when experiencing a lot of pain. Many stated that the use of such elements is for *younger people and kids*, and some of the patients also described having stopped using certain apps *because* of such elements.

Health care providers from collaborating institutions focused more on sharing functionality (ie, possibility of sharing health information with health care providers) than did patients. Despite seeing sharing possibilities as something positive, patients were skeptical as to how this could work and found the option unrealistic given the limited time available for health care providers, and their impressions that health care providers often work nonstop with no availability to respond to email/phone calls during a full workday. When asked what mattered most to them, the patients preferred an intervention that could give them personalized suggestions for exercises and content related to the issues and areas they described as challenging. Many patients also emphasized this as one of the main reasons for wanting simple ways to self-assess or track behaviors (eg, for sleep and activity), wanting the intervention to suggest exercises based on their own personal behavioral patterns.

Each stakeholder workshop also involved a priority task (ie, choosing between different design elements and features), where participants voted (each with a maximum 3 votes in addition to 2 group votes) on potential design elements and features. See Table 5 for details on the distribution of votes from patients (n=15), spouses (n=2), and health care providers (n=2); whether the design element/feature was included in the final app; why/why not; and a few illustrating quotes.

Table 5. Design elements/features: priority task voting, elements/features included, details, and illustrating quotes.

Design element/feature	Description	Votes, n	Included in the final app (yes/no) and details/justification	Illustrating quotes
Customization and personalization	Customize how things are presented/look in the app. For example, you can customize colors, styles, or specific parts of the app that you want to use.	14	Yes. The features My page and My favorites were included to allow for personalization and easy access to preferred content. In addition, the sequence of some of the modules could be individually chosen, to allow for more individual preference.	"For me, it is very important that it is individually tailored/customized."
Behavioral trackers	Map/log what you do to see connections and opportunities for change.	13	Yes. Daily self-assessment/registrations of pain, sleep, rest, activity, and mood were included as optional features for those preferring to track all/some of these factors.	"Today, I've been in a lot of pain, but I don't know why [...] The registrations I'm looking for will tell me why I have so much pain every Thursday."
Feedback	Get feedback from the app. For instance, by telling you what you have achieved lately or show you new ways to do things.	10	Yes. Several of the exercises in the app allow for registration of current habits/activities and give suggestions for new ways to do things.	"I think it should, in a way, replace a personal coach [...] and be able to provide feedback, and discuss with me. What went well, what went wrong."
Automatic tailoring	The app automatically adapts to your personal use. For example, you can bring up content and exercises according to your previous preferences.	10	Yes. The app gives the users suggestions for modules and exercises to try based on their marked favorites.	"You may receive quicker feedback if it is automated, as health care personnel go home at 4 pm."
Visualization	Visualization is used to present content in an engaging and visual way. This can be through the use of animations, cartoons, graphs, etc.	8	Yes. Illustrations and photographs are used in the app to support the content but are presented in a *muted* way so as not to appear overwhelming or challenging. Graphs, illustrating the users' behavior tracking, were also implemented.	"I imagine some pictures of famous places that give me energy, people or animals that give me energy, and nature, that gives me energy."
Communicating with health care professionals	Communicate with health care professionals, for example, by sharing information, asking questions, or receiving feedback.	6	No. Not prioritized because of conflict with the desire for easy access by means of a 4-digit personal identification number, and the desire for an app that can serve as a stand-alone self-management program.	"When you have this kind of an app, it is important that when you push the button, you get right in [without high-level log-in procedures], and especially when you are not feeling good."
Communicate with peers	Communicate with peers/other users, for instance, via forums or share achievements with other users of the app.	2	No. Not prioritized because of potential negative impact, conflict with the desire for easy access, and the notion that this would require a larger user base than planned study inclusion.	"Social contact with other users, I think it can be very negative. You can so easily pull each other down."
Avatar	Create your own avatar, that is, a person you can be/that follows you in the app. You can customize it to look the way you want, for example, by looking like your favorite animal.	0	Yes. On the basis of eHealth expert input and existing research [41,42], the *buddy* EPIOS (a bird) was included as an engaging element to stimulate engagement and adherence.	"It made me think of children when I saw it."
Using metaphors	Metaphors can be used as a motivational way of getting through the program/app. For example, let the app be a garden where you can walk around or groom or plant things.	0	No. As the use of metaphors received no votes and was also considered to be a complicating element for the users, this element/feature was not included.	"I did not vote for it" [metaphors] [because I had only three votes to spend and this feature was not important enough for me].
Rewards and trophies	Points and trophies are collected through using the app. For example, you can go up a level when you have collected enough points or get a trophy for strikes, for example, when you have used the app every day for a week.	0	Yes. On the basis of eHealth expert input and findings from existing research [41,42], rewards and trophies were included as engaging elements to stimulate engagement and adherence.	"I'm not very competitive so it doesn't suit me very well, but I can see that it may be a good thing for others."

Intervention Content Development

Existing interview material [50], workshop discussions and usability testing (presented below) indicated that participants preferred a neutral name for the intervention program, encouraging limited use of negative words, such as *pain*, or *too positive* words, such as *positive focus*. During content development, a project team brainstorming and informal *namecompetition* resulted in the intervention being named EPIO, derived from the Greek mythology goddess Epione, the goddess for the soothing of pain.

As evidence-based psychosocial/educational interventions are mainly conducted in person, the decision was made for the EPIO program to contain a face-to-face introduction session, where participants would receive an introduction to the EPIO intervention program, as well as help in downloading the app onto their smartphone or tablet.

Given the significant evidence of potential for support from CBT and ACT for people with chronic pain[10-18], combined with input from all stakeholders, the EPIO intervention was primarily based on CBT (ie, thought and behavior challenging, cognitive restructuring, behavior change, problem solving, and coping) [6,8] but with aspects of ACT (ie, value-based direction, acceptance of pain, and awareness of the present) [7,9], both resting on aspects found essential for pain management [10,18]. The final app-based EPIO intervention program contained 9 modules, as illustrated in Table 6. Each module in EPIO contained educational topics (eg, about pain, balance and activity pacing, thoughts and feelings, health behaviors, and coping during difficult times), as well as brief topic-related tasks and a variety of relaxation-focused exercises (eg, diaphragmatic breathing, progressive muscle relaxation, visualization, mindfulness, and meditation) anchored in existing treatment manuals and findings for chronic pain management [61,64]. The first 5 modules were presented consecutively in a fixed sequence, as the content in each of these modules was considered essential for the subsequent topics. To allow for more individual preference, the sequence of modules 6 to 8 could then be individually chosen, if preferring to do so.

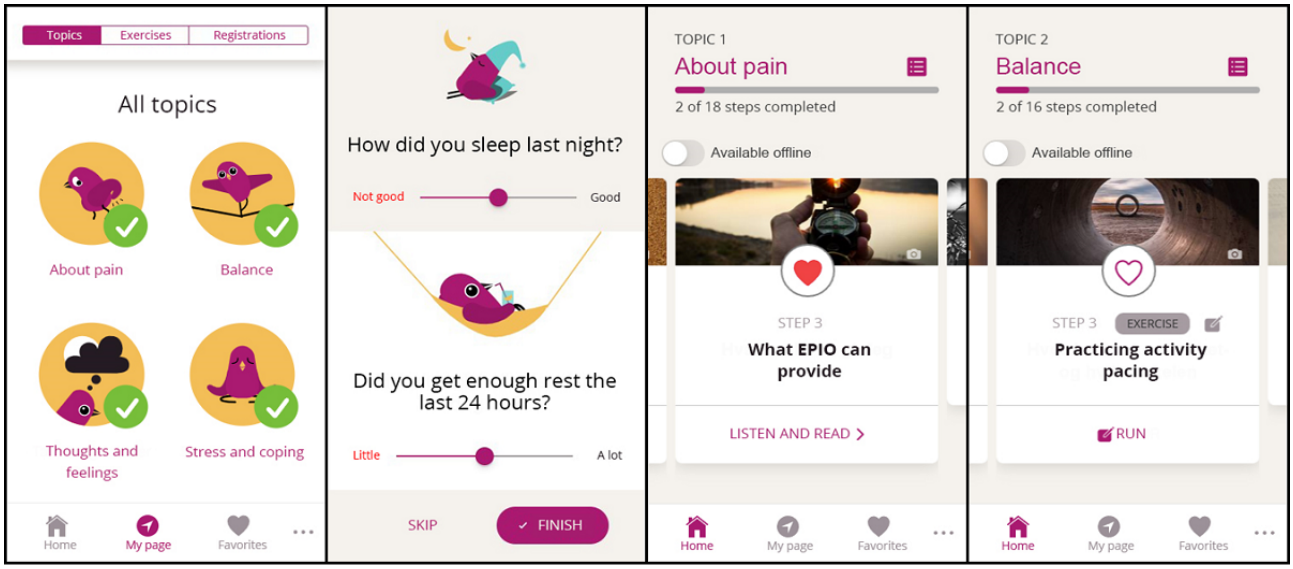
Table 6. Overview of EPIO modules and content.

Module #	Module title	Content
0	Introductory session	60 min in-person/group session. Introduction and intervention overview, practical exercise example, and help in downloading and using the intervention.
1	About pain	Introduction to the intervention program, including information about pain and pain management. Coping strategies, fight-or-flight response, and introduction to breathing and relaxation; rationale and exercises.
2	Balance	Activity pacing and planning, introduction to mindfulness, self-care, pleasant activities, EPIO as your toolbox, and progressive muscle relaxation.
3	Thoughts and feelings	Pain, the relationship between thoughts and feelings, recognizing negative thoughts and cognitive distortions, gratitude, and positive thinking. Exercises including challenging negative thoughts, mindfulness, and autogenic muscle relaxation.
4	Stress and coping	About stress, coping, and rationale for stress management and relaxation strategies. Acceptance, active and passive coping approaches, and visualization.
5	What is important to me?	Defining and exploring individual values and goals. Personal role models, self-image, and intruding thoughts. Planning and goal setting. Introduction to meditation.
6	Health behaviors and lifestyle	Health behaviors and health behavior change. Awareness of important health behaviors, including sleep, physical activity, nutrition, and substance use/abuse; rationale and exercises. Stretch-based relaxation methods/exercises.
7	Communication, relations, and social support	Communication, assertiveness, support systems, and social networks. Exercises related to awareness about social support systems, how to strengthen social support, and progressive muscle relaxation.
8	Coping during difficult times	Self-regulation and implementation of coping strategies in everyday life. Pain, frustration and anger management, daily use of coping strategies in everyday life. Introduction and use of distraction, visualization, and stretch-based relaxation.
9	Summary and the road ahead	Review and summary, where to go from here and advice for the road ahead.

Software Development and Formative Evaluation

Stakeholders' input and 5 iterative rounds of usability testing contributed to adjustments to detect and ensure (1) easy and intuitive navigation, including adding short cuts, introductions, and symbols; (2) language and content issues, including adding more steps to reduce the length of each section, replacing or removing difficult words and terms; (3) implementation of engaging design elements to stimulate adherence, including adding rewards/trophies, as well as an avatar, the *buddy* EPIOS, an animated bird accompanying the users throughout EPIO, and finally d) possibilities for personal preferences and choice, including adding possibilities to choose between reading or listening, and choosing which, if any, variables to track (eg, pain, sleep, activity, mood, and rest). See Figure 5 for screenshot examples.

Figure 5. Example screenshots from the EPIO intervention. From left: (1) start page, (2) selective registrations, (3) module about pain, and (4) exercise example.



The usability testing revealed an overall average SUS score of 81.25, indicating a grade A, which equals to excellent system usability [65]. Users reported that the intervention was easy to use, without need for assistance from anyone. Most users reported that they thought they would use the intervention frequently (mean 3.8, median 4).

In response to input from study participants, a number of adjustments were made for the final version of EPIO. Some of these are described in the following sections.

My Page

As personalization was emphasized as an important feature, a personal page, *My page* was added to the program. This included an overview of the user's personal progress in the intervention program in addition to access to personal registrations, illustrated in graphs. In addition, based on the input from eHealth experts and previous research [41,42], a *trophies* section was included in *My page* to stimulate personalization and motivation. Ability to gauge the length of each step and exercise in seconds/minutes was also implemented.

My Favorites

Patients expressed a need for easy and direct access to exercises and content that they liked; therefore, a *mark as favorite* feature was added to each step in the program. As usability testing revealed usability issues for this feature (ie, difficulties for users to grasp *how* to mark their favorites), the final program version included introductions presented in a step-by-step manner, with the *buddy* bird EPIOS later reminding users of these steps and the option to add the text *add as favorite* at the end of each step. Usability testing revealed that participants liked the bird EPIOS and the brief summaries and reminders provided by EPIOS.

Practicing Mode

CBT typically includes homework between sessions to practice and generalize new skills and behaviors. Therefore, practice and repetition were encouraged in EPIO, and following completion of each module, participants could not open a new module for 3 days. This was done to give users time to practice and implement completed modules, knowledge, and exercises into their daily lives. EPIO provided encouragement for practice, either through suggested steps or through choosing own favorites.

Security and Privacy Considerations

The EPIO intervention program was developed for people with chronic pain. It was developed at and distributed from a major hospital. Protecting patients and patient information is the responsibility of health care providers and the institution (ie, the hospital) and privacy and security were of essence to consider in the design and development process.

One issue concerned the amount of sensitive information and options related to log-in requirements. Participating patients emphasized the importance of a simple log-in procedure. Most of these patients expressed little concern regarding privacy and security protections, stating that it was more important for them to get an accessible tool they could get *direct access to on a bad day*, referring to days with a lot of pain, *without any hassle or things to remember*, such as high-level access procedures and passwords. This was the case in the stakeholder workshops not only in this study but also in previous patient/spouse interviews [50]. At the same time, however, many of the patients wanted to be able to keep personal notes/diaries, and some also wanted to be able to share their information with their health care provider through the app and/or connect with peers using the app. This would introduce further privacy challenges. Adding functionality such as sharing options would increase the privacy level needed and therefore also increase the security requirements. However, privacy and security are essential in these types of settings, and as ease of use and a simple log-in were identified as one of the most important patient requirements, the solution was to incorporate a simple 4-digit PIN, excluding functionality such as sharing possibilities and personal notes. Users were instead encouraged to use a pen and paper and take notes during some of the themes and exercises (ie, "You may find it beneficial to use a pen and paper for this exercise.").

Discussion

Principal Findings

This study describes the design and development process of EPIO, an eHealth pain management intervention for people with chronic pain. The process combining evidence-based and user-centered approaches is a previously recommended but underutilized approach to eHealth intervention development [27,33,34]. To our knowledge, this is the first study combining evidence-based knowledge with stakeholders' input to inform the development of an eHealth intervention for self-management of chronic pain.

The EPIO intervention program was developed using iterative processes through a combination of (1) *contextual inquiry and co-design processes*, where input from people with chronic pain, spouses, health care providers, and other collaborating partners was gathered; (2) *intervention content development*, where relevant content topics were identified and intervention content was created based on clinical experience and with inspiration from existing evidence-based cognitive behavioral and acceptance and commitment pain management programs [61-64]; and (3) *iterative software development and formative evaluation*, including low- and high-fidelity prototypes and usability testing. External stakeholders (ie, patients, spouses, health care providers, and other partners from collaborating institutions) described a number of challenges associated with current options for pain management care, emphasizing the potential within eHealth technology and more available sources for pain management strategies. Patients described the need for an accessible solution that fits within their existing everyday routines, giving them a *break*; positive input; and reminders in their daily lives. To meet acceptability and usability needs for the target group, the intervention used easily understandable language, with brief and to-the-point sections made accessible on small screens and mobile phones. Stakeholders also pointed to a need for intuitive and effective functionalities that did not demand too much of the patients' time, giving them options to choose from and automatic suggestions adjusted to their needs.

Evidence-Based Knowledge and the Importance of User Involvement: Finding the Right Balance

CBT- and ACT-based psychological interventions have been shown to be effective, improving quality of life, depressive symptoms, and pain acceptance for people living with chronic pain [10,11,13,14,16,66,67]. The goal of this study was to design and develop such an intervention to be delivered in a technological format and on a mobile platform. Seeking to achieve persistent change in a person's health and overall well-being, intervention programs must be based on evidence-based knowledge, and according to the Medical Research Council's guidance, all complex interventions should be guided by the latest evidence and appropriate theory [68].

Despite these facts, the development of evidence-based eHealth pain management interventions has been limited [33-35], as has the incorporation of user involvement in these processes. Even evidence-based interventions depend on users' acceptance, adherence, and overall user fit for interventions to be successful [39]. The lack of user involvement (ie, patients and health care providers) and human centeredness in development of eHealth interventions have been criticized repeatedly [5,21,27,32,34,36,39], and this lack of user involvement in the development process can potentially also explain the high attrition rates and low adherence associated with such interventions [32,40-42]. People simply stop using technologies that do not meet their needs, requirements, or daily routines.

In addition to general user requirements (eg, technology being user friendly and flawless), it is essential to incorporate the needs and requirements of the specific user group. For the participants in this study, that meant the eHealth intervention had to accommodate patients' varying and often high pain levels, their challenges with feeling guilty and *never doing enough*, and their concentration issues. Stakeholders stated that it was important that the intervention did not focus on the negative aspects of living with chronic pain or provide users with too much information or too much choice, flashy graphics such as sound and animation effects, or cumbersome log-in procedures. Patients wanted positive input in their daily lives through a solution that provided them with useful, effective, and personalized advice on pain management, reminding them to take smaller breaks during the day. Participants also suggested functionalities to register daily activities, sleep, and mood level so that patients could become more aware of how these areas affected their daily life. Some of the patients also wanted an option to register their daily pain level. From a CBT standpoint, this could be viewed as useful, as increasing patients' awareness and ability to take an active part in one's own life is crucial. However, the literature has shown that too much focus on the pain itself, for instance, through keeping a pain diary, can be negative and could possibly increase pain interference [69-71]. However, studies have also illustrated the positive sides of pain screenings/registrations. As several participants regarded self-assessment/registrations as important, and this was also voted high on the prioritizing task (Table 5), self-assessment/registrations were included in the initial development.

The design and development process in this study did reveal some disagreements between what was considered important by health care providers and other collaborating partners versus what was considered important by some of the patients. Although health care providers emphasized the need for available, evidence-based, and trustworthy information given to patients, seeing eHealth technology as a positive option for providing patients with such knowledge, patients expressed some conflicting views on the topic. Patients generally agreed that information and content should be trustworthy, yet they kept emphasizing during workshops as well as usability testing that they did not want *too much* information, that they *already knew* a lot about pain and the theory behind pain management, and that they first and foremost wanted effective and quick exercises that could help in their daily lives. This could have been because several patients were recruited from patient education centers and courses focusing on chronic pain, and thus, they already had received a lot of information. However, literature has shown that compared with the general population, people living with long-term conditions report more difficulties

with understanding health information in addition to having greater difficulties in engaging with health care providers [72]. This, together with some chronic pain patients' reported concentration issues, illustrates the importance of providing chronic pain patients with easily accessible information.

Interventions, and perhaps particularly eHealth interventions, have the challenges of enhancing motivation for use, adherence to use, and motivation for continued use. How to best present evidence-based content is a question of user involvement, acceptability, usability, and feasibility. To promote user engagement and continued use, the EPIO intervention had to present the material in a way that met the users' interests and requirements. Participating health care providers and other collaborating partners with eHealth expertise, as well as software developers, suggested adding gamifying design elements such as rewards and avatars, emphasizing the importance of engaging and motivational design elements and pointing to evidence that shows that the use of such elements and persuasive technology positively affects adherence and well-being [41,42]. However, the participating patients found this less important, and they also expressed concern that the use of such effects could be potentially challenging when in pain. None of the patients voted for such elements in the priority task in the stakeholder workshops (see Table 5). Patients instead stated that they wanted the content presented in a simple way.

On the basis of these findings, it was important to find a balance in the use of design elements, with the final EPIO program including some of these types of elements, such as trophies for progress and continued use; and an avatar/buddy, the bird EPIOS; and providing users with content summaries and brief motivational messages. The buddy bird EPIOS, therefore, has an educational role in the intervention program, in line with what users emphasized as important, but at the same time, EPIOS has a motivational and relational role, in line with participating health care providers and eHealth experts as well as recommendations from existing literature [42].

Strengths, Limitations, and Future Directions

This study presents some limitations that need to be considered. First, the limited number of male and younger patients (mean age 51 years) might limit the representativeness of the study. Given the large percentage of female patients compared with male patients participating in this study, despite encouraging participation of both genders living with chronic pain, the patient sample can be considered a sample of convenience. However, it should be noted that the prevalence of chronic pain is higher among females compared with males [73], and research also shows challenges in recruiting male participants compared with female participants for intervention programs focusing on self-management [74]. In addition, the participating patients represented a wide range of pain diagnoses, and as chronic pain is more prevalent among people older than 50 years [75], it may be argued that the patients participating in this study were in fact representative of the user group. The fact that some of the patients and spouses had also participated in an initial interview study [50] could potentially also be a limitation, as it is possible that other opinions and perspectives would have emerged if more novel users without prior knowledge of the emerging intervention had participated. However, mutual learning and shared understanding are core concepts within participatory design, as this is the only way to ensure mutual respect between stakeholders, enabling everyone to take part in the shared decision-making process [56]. Patients are not eHealth experts and do not necessarily have the language to articulate what they need from an eHealth intervention. Consequently, using the same sample of participants and giving them enough knowledge about design and development processes may have made it easier for the participating patients to take an active part in development discussions. However, this study did also include new and *naïve* patients with chronic pain to add to previously collected qualitative data.

The software development and formative evaluation may also present some limitations. For instance, every part of the intervention steps/modules was not tested. However, the intervention material was based on the same concepts and foundations, written by the same experienced team [76,77], led by the same PI, and using the same therapeutic language and structure. Therefore, it was considered more important to get users' feedback on functionality and design, including layouts and how the material was presented, than on every written word. Usability testing was conducted at the project team offices, with a facilitator and observer watching the participants within a limited period. This could have impacted the testing, and it is possible that the usability testing of EPIO captured only some of the potential barriers to continued use over time.

A number of strengths are also evident in the current design and development process. As recommended by existing research [5,27,39] and to ensure trustworthiness [78], the study included a broad range of stakeholders (ie, patients, spouses, health care providers, and eHealth experts as well as researchers, editors, software developers, and user representatives) from the project planning stage and throughout the development process.

The development process was guided by existing development recommendations, using a broad range of service design methods and a user-centered design approach to facilitate cocreation, mutual learning, and shared understanding among the stakeholders involved. Intervention acceptability (ie, well received, suitable, user friendly, attractive, and meeting needs) to users was one of the main goals for the design and development of EPIO. Although the intervention program was developed using a participatory design approach to support the likelihood of acceptability, usability, and feasibility, acceptability will need to be further tested and established in a future pilot test study before instigating efficacy studies. Given the challenge of low adherence and high attrition rates in eHealth interventions [41,42], the development of the current intervention sought to incorporate stakeholder-identified aspects supporting adherence. Whether this turns out to be an effective approach facilitating adherence needs to be evaluated in a future pilot test and subsequent randomized controlled trial (RCT). In addition to high user involvement and stakeholder input, the development process was guided by theory and concepts from well-established cognitive behavioral and acceptance and commitment pain management programs, meeting the requests for eHealth pain management interventions that are based on evidence-based knowledge. This enhances the future potential

for positive effect findings. The ultimate goal of the EPIO program is to support improvements in quality of life for people with chronic pain. Therefore, in addition to test usability, feasibility, and acceptability, the next step in the research process will be to examine preliminary efficacy findings in a pilot test before eventually examining the intervention in a full-scale RCT. Together with the high focus on privacy and security aspects, acceptability and efficacy are also likely to increase the potential for poststudy implementation.

Conclusions

This study offers insight into how to take a user-centered approach to the design and development of an evidence-based eHealth pain management intervention for people with chronic pain. Developing evidence-based eHealth interventions while also involving the voices and perspectives of a variety of stakeholders can be challenging, time consuming, and sometimes even an expensive process. However, continuing to develop and test non–evidence-based, non–user-centered interventions is not a great alternative. Instead, mutual learning and shared understanding become crucial. This study involved patients, spouses, health care providers, and other relevant stakeholders in the design and development process of the eHealth interventions, pointing to important steps for developing useful and meaningful interventions for patients. In addition to informing the potential process of developing an eHealth pain management intervention, this study also provides a practical example of how eHealth interventions can be designed and developed to combine evidence-based material with user-centered requirements. To test usability, acceptability, and feasibility, as well as the potential efficacy of the program, further research is needed and a pilot test is currently underway to optimize the EPIO program in preparation for a full-scale RCT.

Acknowledgments

This study was supported by the Norwegian Research Council, grant number 256574 (PI: LSN). The authors would like to thank the patients, spouses, and health care providers who participated in this study. The authors would also like to thank all collaborating project team partners, including the editorial group and software team at Department of Digital Health Research at Oslo University Hospital, particularly editorial team leader Marianne Westeng, editor Hilde Renate Sætre Sleveland, editor Elanor Halvorsen Brendmo, designer Yizhak Itzchaki, technical lead Stein Jakob Nordbø, and IT developer Per Olav Flaten for their exceptional work throughout the project process. The authors would also like to thank researcher Jelena Mirkovic and former PhD student Karina Heldal for their input in the project.

Conflicts of Interest

None declared.

References

1. 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]
2. Melzack R. From the gate to the neuromatrix. Pain 1999 Aug(Suppl 6):S121-S126. [doi: 10.1016/s0304-3959(99)00145-1] [Medline: 10491980]
3. Nes LS, 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]
4. Nes LS, 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]
5. Kress HG, Aldington D, Alon E, Coaccioli S, Collett B, Coluzzi F, et al. A holistic approach to chronic pain management that involves all stakeholders: change is needed. Curr Med Res Opin 2015;31(9):1743-1754. [doi: 10.1185/03007995.2015.1072088] [Medline: 26172982]
6. Beck AT. Cognitive Therapy And The Emotional Disorders. New York: Penguin Books, Limited; 1991.
7. Hayes SC, Strosahl KD, Wilson KG. Acceptance And Commitment Therapy: An Experiential Approach To Behavior Change. New York: The Guilford Press; 1999.
8. Beck JS. Cognitive Behavior Therapy, Second Edition: Basics And Beyond. New York: The Guilford Press; 2011.
9. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies - republished article. Behav Ther 2016 Nov;47(6):869-885. [doi: 10.1016/j.beth.2016.11.006] [Medline: 27993338]
10. Craner JR, Sperry JA, Koball AM, Morrison EJ, Gilliam WP. Unique contributions of acceptance and catastrophizing on chronic pain adaptation. Int J Behav Med 2017 Aug;24(4):542-551. [doi: 10.1007/s12529-017-9646-3] [Medline: 28299623]
11. Eccleston C, Morley SJ, Williams AC. Psychological approaches to chronic pain management: evidence and challenges. Br J Anaesth 2013 Jul;111(1):59-63 [FREE Full text] [doi: 10.1093/bja/aet207] [Medline: 23794646]
12. Gilliam WP, Craner JR, Morrison EJ, Sperry JA. The mediating effects of the different dimensions of pain catastrophizing on outcomes in an interdisciplinary pain rehabilitation program. Clin J Pain 2017 May;33(5):443-451. [doi: 10.1097/AJP.0000000000000419] [Medline: 27437567]
13. 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]

14. McCracken LM, Turk DC. Behavioral and cognitive-behavioral treatment for chronic pain: outcome, predictors of outcome, and treatment process. Spine (Phila Pa 1976) 2002 Nov 15;27(22):2564-2573. [doi: 10.1097/00007632-200211150-00033] [Medline: 12435995]
15. McCracken LM, Vowles KE. Acceptance and commitment therapy and mindfulness for chronic pain: model, process, and progress. Am Psychol 2014;69(2):178-187. [doi: 10.1037/a0035623] [Medline: 24547803]
16. 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. [doi: 10.1016/s0304-3959(98)00255-3] [Medline: 10204712]
17. Nash VR, Ponto J, Townsend C, Nelson P, Bretz MN. Cognitive behavioral therapy, self-efficacy, and depression in persons with chronic pain. Pain Manag Nurs 2013 Dec;14(4):e236-e243. [doi: 10.1016/j.pmn.2012.02.006] [Medline: 23273826]
18. Pots WT, Trompetter HR, Schreurs KM, Bohlmeijer ET. How and for whom does web-based acceptance and commitment therapy work? Mediation and moderation analyses of web-based ACT for depressive symptoms. BMC Psychiatry 2016 May 23;16:158 [FREE Full text] [doi: 10.1186/s12888-016-0841-6] [Medline: 27215733]
19. Andersson G, Ljótsson B, Weise C. Internet-delivered treatment to promote health. Curr Opin Psychiatry 2011 Mar;24(2):168-172. [doi: 10.1097/YCO.0b013e3283438028] [Medline: 21285706]
20. 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]
21. 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]
22. 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]
23. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr;42(2):169-187. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
24. 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]
25. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
26. Andersson G. The Internet and CBT: A Clinical Guide. Abingdon, United Kingdom: Routledge, Taylor & Francis group; 2014.
27. 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]
28. 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]
29. 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]
30. 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]
31. Vugts MA, Joosen MC, van der Geer JE, Zedlitz AM, Vrijhoef HJ. 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]
32. 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 Jan 20;19(1):e26 [FREE Full text] [doi: 10.2196/jmir.5951] [Medline: 28108429]
33. 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]
34. 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]
35. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. Br J Pain 2016 Aug;10(3):135-140 [FREE Full text] [doi: 10.1177/2049463716638700] [Medline: 27583140]
36. 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]

37. 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]
38. Moen A, Hackl WO, Hofdijk J, van Gemert-Pijnen L, Ammenwerth E, Nykänen P, et al. eHealth in Europe - status and challenges. Yearb Med Inform 2013;8:59-63. [Medline: 23974549]
39. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
40. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014 Feb 26(2):CD010152 [FREE Full text] [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
41. 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]
42. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. 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]
43. Rosser BA, Vowles KE, Keogh E, Eccleston C, Mountain GA. Technologically-assisted behaviour change: a systematic review of studies of novel technologies for the management of chronic illness. J Telemed Telecare 2009;15(7):327-338. [doi: 10.1258/jtt.2009.090116] [Medline: 19815901]
44. McCormack B, van Dulmen S, Eide H, Skovdahl K, Eide T, editors. Person-Centred Healthcare Research. Hoboken: Wiley Blackwell; 2017.
45. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
46. 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 1;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
47. Rothgangel A, Braun S, Smeets R, Beurskens A. Design and development of a telerehabilitation platform for patients with phantom limb pain: a user-centered approach. JMIR Rehabil Assist Technol 2017 Feb 15;4(1):e2 [FREE Full text] [doi: 10.2196/rehab.6761] [Medline: 28582249]
48. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
49. Varsi C, Nes LS, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
50. Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for eHealth pain management interventions: qualitative study. J Med Internet Res 2019 Apr 1;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
51. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
52. Abras C, Maloney-Krichmea D, Preece J. User-centered design. In: Bainbridge WS, editor. Berkshire Encyclopedia Of Human-computer Interaction: When Science fiction becomes science fact. Massachusetts: Berkshire Publishing Group; 2004:757-768.
53. 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]
54. Hooten WM. Chronic pain and mental health disorders: shared neural mechanisms, epidemiology, and treatment. Mayo Clin Proc 2016 Jul;91(7):955-970. [doi: 10.1016/j.mayocp.2016.04.029] [Medline: 27344405]
55. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
56. Simonsen J, Robertson T. Routledge International Handbook of Participatory Design. Abingdon, United Kingdom: Routledge; 2013.
57. Stickdorn M, Schneider J. This is Service Design Thinking: Basics, Tools, Cases. Hoboken, New Jersey, United States: Wiley; 2012.
58. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 1;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]

59. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
60. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
61. Mayo Clinic. Mayo Clinic Guide to Pain Relief. Second Edition. Rochester: Mayo Clinic; 2014.
62. Stanos S. Focused review of interdisciplinary pain rehabilitation programs for chronic pain management. Curr Pain Headache Rep 2012 Apr;16(2):147-152. [doi: 10.1007/s11916-012-0252-4] [Medline: 22427179]
63. Veehof MM, Schreurs K, Hulsbergen M, Bohlmeijer E. Leven met pijn. De kunst van het aanvaarden. Amsterdam: Boom; 2012.
64. Waxenberg LB. Pain Management Manual & Workbook: A guide for Behavioral Pain Management Techniques. Gainesville: Department of Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida; 2008.
65. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, California , US: Createspace Independent Pub; 2011.
66. Hughes LS, Clark J, Colclough JA, Dale E, McMillan D. Acceptance and Commitment Therapy (ACT) for chronic pain: a systematic review and meta-analyses. Clin J Pain 2017 Jun;33(6):552-568. [doi: 10.1097/AJP.0000000000000425] [Medline: 27479642]
67. Veehof MM, Oskam M, Schreurs KM, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542 [FREE Full text] [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
68. 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]
69. Ahluwalia SC, Giannitrapani KF, Dobscha SK, Cromer R, Lorenz KA. 'It Encourages Them to Complain': a qualitative study of the unintended consequences of assessing patient-reported pain. J Pain 2018 May;19(5):562-568. [doi: 10.1016/j.jpain.2017.12.270] [Medline: 29421247]
70. Arntz A, Dreessen L, Merckelbach H. Attention, not anxiety, influences pain. Behav Res Ther 1991;29(1):41-50. [doi: 10.1016/s0005-7967(09)80006-5] [Medline: 2012588]
71. Vlaeyen JW, Morley S, Crombez G. The experimental analysis of the interruptive, interfering, and identity-distorting effects of chronic pain. Behav Res Ther 2016 Nov;86:23-34. [doi: 10.1016/j.brat.2016.08.016] [Medline: 27614948]
72. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016 Jan 14;6(1):e009627 [FREE Full text] [doi: 10.1136/bmjopen-2015-009627] [Medline: 26769783]
73. Tsang A, von Korff M, Lee S, Alonso J, Karam E, Angermeyer MC, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. J Pain 2008 Oct;9(10):883-891. [doi: 10.1016/j.jpain.2008.05.005] [Medline: 18602869]
74. Gavarkovs AG, Burke SM, Reilly KC, Petrella RJ. Barriers to recruiting men into chronic disease prevention and management programs in rural areas: perspectives of program delivery staff. Am J Mens Health 2016 Nov;10(6):NP155-NP157. [doi: 10.1177/1557988315596226] [Medline: 26206160]
75. Granan LP, Reme SE, Jacobsen HB, Stubhaug A, Ljoså TM. The Oslo University Hospital Pain Registry: development of a digital chronic pain registry and baseline data from 1,712 patients. Scand J Pain 2019 Apr 24;19(2):365-373. [doi: 10.1515/sjpain-2017-0160] [Medline: 30699072]
76. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018 Sep 6;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
77. Børøsund E, Varsi C, Clark MM, Ehlers SL, Andrykowski MA, Sleveland HR, et al. Pilot testing an app-based stress management intervention for cancer survivors. Transl Behav Med 2019 Apr 24. [doi: 10.1093/tbm/ibz062] [Medline: 31330023]
78. Lincoln YS, Guba EG. Naturalistic Inquiry. Beverly Hills, California: Sage; 1985.

Abbreviations

ACT: acceptance and commitment therapy
AES: Advanced Encryption Standard
CBT: cognitive behavioral therapy
eHealth: electronic health
IT: information technology
PI: principal investigator
RCT: randomized controlled trial

SUS: System Usability Scale

Edited by G Eysenbach; submitted 10.09.19; peer-reviewed by T Janmohamed, A Schäfer; comments to author 01.10.19; revised version received 18.10.19; accepted 16.12.19; published 21.01.20.

Please cite as:
Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KMG, Waxenberg LB, Weiss KE, Morrison EJ, Haaland-Øverby M, Bevan K, Zangi HA, Stubhaug A, Solberg Nes L
A User-Centered Approach to an Evidence-Based Electronic Health Pain Management Intervention for People With Chronic Pain: Design and Development of EPIO
J Med Internet Res 2020;22(1):e15889
URL: http://www.jmir.org/2020/1/e15889/
doi:10.2196/15889
PMID:31961331

Original Paper

A Data-Driven Social Network Intervention for Improving Organ Donation Awareness Among Minorities: Analysis and Optimization of a Cross-Sectional Study

Michael Douglas Murphy[1*], BSc; Diego Pinheiro[2*], PhD; Rahul Iyengar[1], BSc; Gene Lim[3], BSc; Ronaldo Menezes[4], PhD; Martin Cadeiras[2], MD

[1]Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[2]Department of Internal Medicine, University of California, Davis, Sacramento, CA, United States

[3]Mav12 Inc, Santa Monica, CA, United States

[4]Department of Computer Science, University of Exeter, Exeter, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Michael Douglas Murphy, BSc
Department of Medicine
University of California, Los Angeles
CHS Building, 1st Fl
Los Angeles, CA,
United States
Phone: 1 6263406288
Email: mikedmurph@gmail.com

Abstract

Background: Increasing the number of organ donors may enhance organ transplantation, and past health interventions have shown the potential to generate both large-scale and sustainable changes, particularly among minorities.

Objective: This study aimed to propose a conceptual data-driven framework that tracks digital markers of public organ donation awareness using Twitter and delivers an optimized social network intervention (SNI) to targeted audiences using Facebook.

Methods: We monitored digital markers of organ donation awareness across the United States over a 1-year period using Twitter and examined their association with organ donation registration. We delivered this SNI on Facebook with and without optimized awareness content (ie, educational content with a weblink to an online donor registration website) to low-income Hispanics in Los Angeles over a 1-month period and measured the daily number of impressions (ie, exposure to information) and clicks (ie, engagement) among the target audience.

Results: Digital markers of organ donation awareness on Twitter are associated with donation registration (beta=.0032; $P<.001$) such that 10 additional organ-related tweets are associated with a 3.20% (33,933/1,060,403) increase in the number of organ donor registrations at the city level. In addition, our SNI on Facebook effectively reached 1 million users, and the use of optimization significantly increased the rate of clicks per impression (beta=.0213; $P<.004$).

Conclusions: Our framework can provide a real-time characterization of organ donation awareness while effectively delivering tailored interventions to minority communities. It can complement past approaches to create large-scale, sustainable interventions that are capable of raising awareness and effectively mitigate disparities in organ donation.

(J Med Internet Res 2020;22(1):e14605) doi:10.2196/14605

KEYWORDS

organ donation; social media; minority health; community health education

Introduction

Background

Organ transplantation is the therapy of choice for patients with end-stage organ failure. Over the past three decades, organ transplantation has saved more than 2 million life-years in the United States alone [1]. However, only half of US adults are registered as organ donors [2], and the current pool of recovered organs inadequately meets the particular medical demand of patients from ethnoracial minorities [3]. With the current shortage of organ donors and an ever-increasing incidence of end-stage organ failure, the number of patients left in need of organ transplantation has grown: only 1 out of 4 patients on the organ wait list will eventually receive the organ transplant needed [4,5]. The success of organ transplantation depends on the patient's histocompatibility with the donated organ, which reaches higher similarities with donors from comparable ethnoracial communities [6,7]. However, the current pool of available organs mainly consists of organs from nonminority donors because of the disproportionate scarcity of ethnoracial minority donors [8]. Increasing the general number of organ donors can mitigate the overall organ shortage, but we can only effectively address the disproportional need of patients from underrepresented demographics by specifically increasing the number of ethnoracial minority donors.

The lack of minority organ donors is generally attributed to insufficient health literacy, which affects how individuals make educated health decisions about their lives and the lives of their families and overall community [9-12]. In the case of organ donation, health literacy specifically impacts the likelihood of individuals to register as organ donors and to consent for the organ donation of their relatives [13,14]. Given that individuals from minority communities tend to have lower health literacy than their ethnoracial majority counterparts, these communities have a relatively lower likelihood of registering as organ donors [9,15,16]. To effectively address this disparity, we need to raise awareness among individuals from minority communities by supplementing them with tailored educational materials about organ donation.

Educational interventions such as the National Minority Organ Tissue Transplant Education Program have generated large-scale and sustainable change across minority communities by raising health literacy [3,14,17-19]. Sustainable large-scale diffusion of health education mainly depends on individual willingness to disseminate the health education received within their social network and how well an individual's social network is integrated within the relevant social constructs as a whole [20,21]. Individuals are more willing to disseminate educational content that is socioculturally tailored and content that is being already disseminated via existing social ties, including family, friends, and other individuals within their community [22,23]. To reach and increase the willingness of individuals from minority communities, health care professionals have created community-based interventions by targeting individuals within these communities with educational content that is socioculturally tailored [3,13,14]. Naturally, a community-based intervention indirectly targets individuals who are likely to be socially connected and, thus, can independently reinforce the dissemination of educational content through social ties. Therefore, community-based interventions not only reach individuals from minority communities but potentiate sustainable change. However, minority communities are not just unintegrated from the whole social system but also isolated from each other, making these traditional interventions ineffective in diffusing health education among these communities at a large scale [3,14].

Objectives

Web-based social networking platforms, also known as social media (eg, Twitter and Facebook), have been proposed as modern venues for the cost-effective delivery of large-scale health interventions with higher outreach in domains as diverse as physical activities, smoking cessation, weight loss, and mental health [20-25]. As social media can be a proxy for real social networks [26], social media platforms are exceptionally suitable for health interventions in which the implicated spreading phenomena are mainly driven by social mechanisms [10,27-29] and can facilitate the delivery of *network interventions* [23]. Network interventions foster higher cascades of behavioral health changes by leveraging the network structure underlying the social context of targeted individuals [30,31]. For instance, the simple decision to register for an internet-based health forum can involve a complex contagion in which individuals require independent social reinforcement and are more susceptible to change their behavior as more peers change theirs [22].

Previous studies have demonstrated the potential of social media to enhance organ donation by promoting health awareness and increasing the number of donor registration rates among minorities [13,32]. However, we still lack a comprehensive framework that allows us to effectively monitor and deliver large-scale network-based interventions of health literacy in real time. We proposed a data-driven framework for improving organ donation awareness by monitoring awareness regarding organ donation and delivering an optimized social network intervention (SNI) using 2 distinct social media interfaces: Twitter for monitoring and Facebook for intervention. Using our framework, we monitored awareness about organ donation over 1 year, then developed and implemented an SNI for improving awareness among minorities over 1 month. The results suggested that our framework can provide a real-time characterization of awareness about organ donation while optimizing the delivery of SNIs to individuals from minority communities. Our data-driven framework has the potential to effectively create large-scale and sustainable interventions to improve organ donation awareness among minorities.

Methods

Identification of Structural Disparities in Organ Donation

To structurally assess disparity, we modeled the connectivity between organ donors and transplant recipients with geographical social networks (GSNs) using the United Network for Organ Sharing (UNOS) database [29]. This dataset includes approximately 438,000 organ transplants conducted in the

United States between 1987 and 2010 containing clinical, geographic, and social information about donors and recipients. In our GSN, nodes are home locations of donors or recipients at the zip code level, and links are organ transplants that were recovered from organ donors living at the origin node and transplanted into recipients living at the destination node. We built a separate ethnoracial GSN for Hispanics, blacks, and whites, focused on recipients [29]. For instance, in the white GSN, the destination node of every link is the home address of a white recipient, whereas the origin node can be the home address of donors from any race/ethnicity. Note that origin nodes (ie, home address of donors) can also be destination nodes (ie, home address of recipients). Finally, we had 3 ethnoracial GSNs that represent the structure of the organ transplantation flow for each race/ethnicity.

Using network science [31-34], we compared our GSNs by quantifying the local and global connectivity according to GSN-respective clustering coefficients and the average path lengths. The clustering coefficient () quantifies the likelihood of 2 nodes being connected, given they share a common node, ranging from 0 (ie, low clustering) to 1 (ie, high clustering). For instance, in a social network of friendships, a clustering coefficient can quantify how likely my friends are also friends with one another. In our GSN, this measure quantifies how likely organ transplants occur between home addresses A and B given that they occur from home address C to both home addresses A and B. This measure of clustering between nodes within a single local network is an influential factor in ascertaining network shortcomings or structural disparities, which could lead to unequal access to donor organs.

Similarly, the average path length (L) is a global measure of connectivity, and it quantifies the typical number of links connecting 2 nodes in the whole network, ranging from 1 to the diameter of the network (ie, the shortest to longest path length between 2 nodes). In a social network of friendships, for instance, the average path length quantifies how many friends typically separate 2 individuals. In our GSN, the average path length quantifies the number of links that typically separate any 2 home addresses among which organ transplants are occurring. This measure of relative accessibility among connected nodes within a global network is an influential factor in uncovering structural disparities, which could lead to strained or unsuitable access to donor organs [31-33].

Finally, we also identified the communities of home addresses with similar organ transplantation dynamics within each ethnoracial GSN using community detection [34,35]. For each network, we measured the number of nodes (N_n), links (M), average degree (M/N_n), clustering coefficient (CC), average path length (L), and the number of communities (N_c). Owing to the underlying network of organ transplantation flow, the connectivity measures along with the number of communities attempt to assess the structural disparity in organ transplantation.

Digital Sensor for Organ Donation Awareness in Social Media

In past work, we have explored the extent to which social media (ie, Twitter) can be used as a sensor for organ donation awareness [28,36]. Twitter is a convenient tool for real-time social sensing because it allows for data collection from most of its users as long as these users set their profile as public. We demonstrated that Twitter has sufficient information regarding organ donation awareness and has the potential to be employed as a social sensor for organ donation campaigns by characterizing conversations according to the volume of mention to different solid organs [28,36].

The organ-related tweets were automatically collected using the minimalist Twitter application programming interface (API) for Python [37], which searches the Twitter stream API, constraining the search by filtering the tweets containing a predefined set of organ donation digital markers among the 140 characters of the tweet text. Organ donation digital markers were defined based on a set of 5 context words (ie, transplant, transplantation, donor, donation, and donate) and a set of 6 subject words (ie, heart, kidney, liver, lung, pancreas, and intestine). For the subject words, only the 6 major solid organs were included, and other possible subject words such as cornea, bone, and skin were not considered. This approach ensures that each collected tweet contains at least one of the 5 words from the context set and at least one of the 6 words from the subject set. Besides, it also ensures that the individuals who wrote these tweets are aware of at least one aspect of organ donation.

Each collected tweet was subsequently augmented with its user's location. Only 0.49% (4875/975,021) of tweets contained the global positioning system coordinates from where the tweet was posted. Therefore, a structural address containing the country, state, county, city, and zip code was automatically extracted from the self-reported location contained in the user profile using the python package geopy and the Nominatim search engine for OpenStreetMap data [38]. Finally, augmented tweets were filtered to only retain those belonging to US users. Therefore, our final tweet dataset was conceived in the context of organ donation and included 1 year of data representing more than 70,000 users in the United States.

Calibration and Efficacy of the Digital Sensor

To validate the extent to which the organ-related tweets collected using Twitter could be used as a digital sensor for organ donation awareness in social media, we assessed the association between the number of organ-related tweets collected by the digital sensor and the number of organ donor registrations. The data of organ donor registrations were obtained from Donate Life California [39]. It contains donor registrations at the zip code level from Los Angeles county. Owing to the scarcity of tweet data at the zip code level, the number of organ-related tweets and donor registrations were both subsequently aggregated at the city level. Afterward, a Poisson regression model was used to model the number of organ donor registrations as a function of the number of organ-related tweets and the size of the population at the city level. A data-intensive approach was used as a second independent model for validating the consistency of the Poisson regression model. The data-intensive approach grouped cities into 4 groups of incremental tweet rate percentile intervals: 0-25, 25-50, 50-75, and 75-100. Afterward, for each group, it estimates the organ

donor registration rate using 10,000 bootstrap samples with replacement.

Digital Intervention Using the Facebook Advertising Platform

Our intervention consisted of targeting Facebook users with educational materials about organ donation via Facebook's advertising platform. Our content comprised short motivational videos associated with testimonials, current facts, and statistics about organ donation, as well as a link to the organ donation registration website (Donate Life California, Sacramento, California) [40]. All text and content used as educational content for the intervention was developed in collaboration with One Legacy, an organ procurement organization (OPO) for Southern California. OPOs follow the best practices in the development of material for organ donation, which is guided by diverse and multidisciplinary focused groups.

Using Facebook's advertising platform from August 4 to September 3, 2016, we systematically targeted communities found to be at risk for a structural disparity. The criteria were based on location, sex, age, and income level, and thus, the intervention was delivered to a selected audience instead of a mass of incidental recipients. Targeting implicated individuals, such as in community-based interventions, can improve the intervention effectiveness because it increases the likelihood of targeting connected individuals who in turn are more likely to act as social reinforcers for others [41]. Targeting these connected individuals also facilitates the creation of organic sustainability by the mechanisms of engagement existing on Facebook (eg, like and share) [41,42]. After our intervention initially exposes educational content to targeted users on Facebook, these users can actively disseminate the targeted content among their social network and, thus, contribute to the exposure of these contents to other individuals who were not previously targeted by the intervention in the first place [20]. This additional organic exposure is ultimately controlled by Facebook's algorithm, which is inherently biased toward targeting these exposures to similar users.

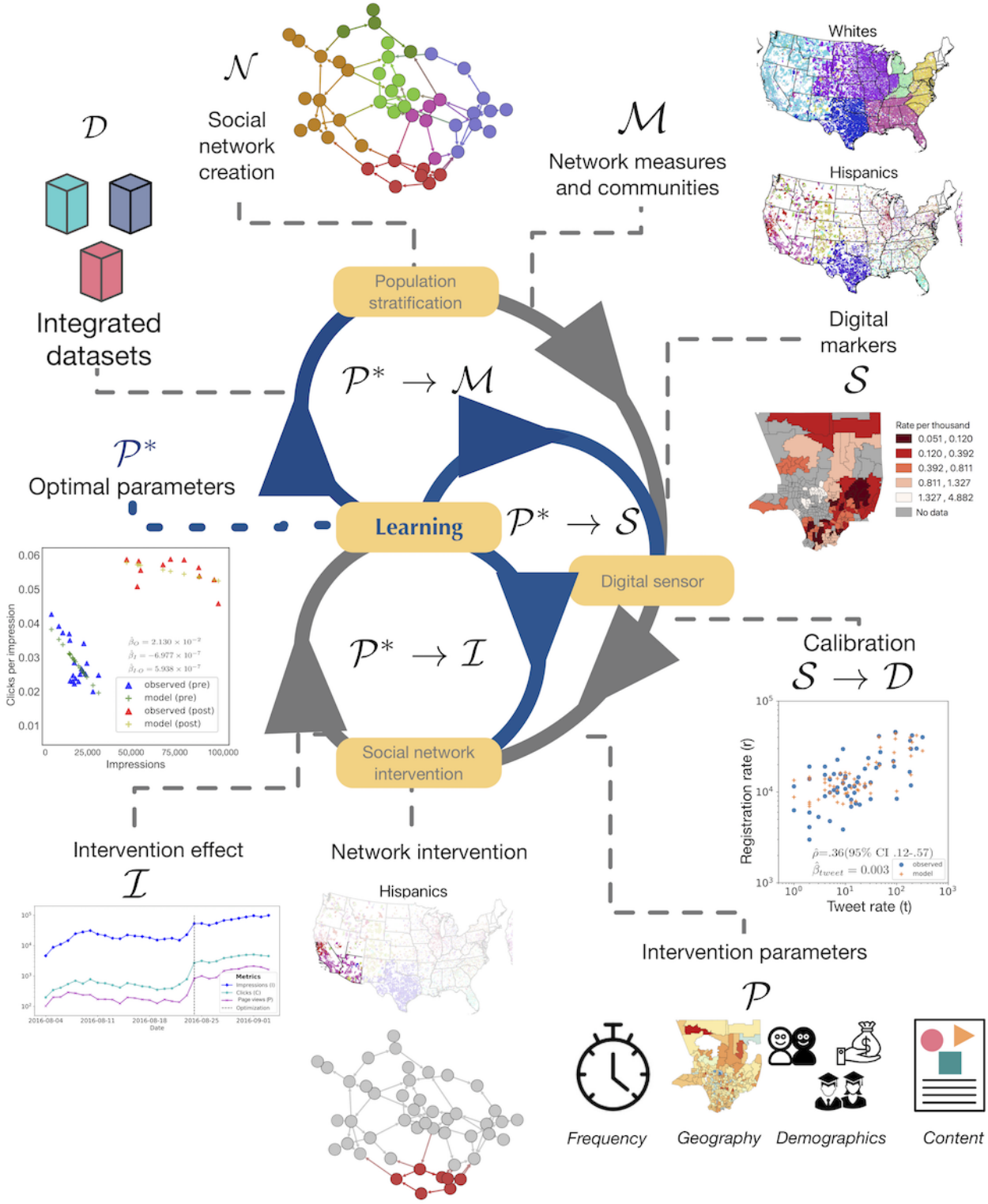
Measure Effect and Optimization

The number of impressions (*I*), clicks (*C*), and page views (*V*) were used to measure the effectiveness of our SNI. These measurements were provided daily by the Facebook advertising platform throughout the intervention. Our SNI delivered content in 2 phases: pre optimization and post optimization. In the preoptimization phase, from August 4 to August 23, the SNI delivered all content with equal proportion and calculated the number of clicks per impression (*C/I*) associated with each content. At the end of the preoptimization period, the SNI learned which content had the highest capability of fostering active engagement among the target audience as measured by the content's *C/I* ratio. Afterward, in the postoptimization phase, from August 24 to September 3, the intervention was optimized to deliver the educational content associated with the highest *C/I* ratio. Given the absence of a baseline, we used the optimization as an instrumental variable and considered the intervention before optimization as a control group for the intervention after optimization. Ordinary least squares (OLS) regression was used to model the number of clicks per impression (*C/I*) as a function of both the number of impressions (*I*) and the use of optimization (*O*). The optimization was outsourced to the company MAV 12.

The overall framework of the SNI is summarized in Figure 1. To characterize structural disparity, separate ethnoracial network-based community analysis was performed. To characterize population awareness about organ donation, data mining of the digital markers of organ donation awareness was performed using Twitter and subsequently calibrated using organ donation registrations. Educational content was delivered to the targeted audience using Facebook in 2 phases: preoptimization and postoptimization. In the preoptimization phase, the SNI delivered all contents and calculated the number of clicks per impression associated with each content. At the end of the preoptimization period, the SNI learned which content had the highest rate of clicks per impression. Afterward, in the postoptimization phase, the intervention was optimized to deliver the educational content associated with the highest rate of clicks per impression.

In general, the SNI characterizes the communities within the transplantation system using network analysis and monitoring the digital markers of organ donation awareness using Twitter. The calibration of these markers on Twitter was performed in conjunction with existing datasets of donation registration from Donate Life, which substantiated the delivery of a large-scale SNI using Facebook. Real-time data were collected to uncover the optimal content for user engagement, which allowed us to optimize the intervention to better target the intended demographics. The University of California Los Angeles Investigational Review Board approved the study.

Figure 1. Conceptual framework of the optimized social network intervention.



Results

Assessment of Disparities in Organ Donation

Each of the ethnoracial GSNs focused on organ transplant recipients elucidates both local and global measures of connectivity as well as the varying number of ethnoracial communities within the whole social system (Table 1). The Hispanic GSN has an average degree (M/N_n) that indicates that Hispanic recipients typically receive organs from a fewer number of distinct donor addresses. Furthermore, the Hispanic GSN had the highest average path length (L), which indicates that Hispanic recipients receive organs from donors living further away in their social network. In addition, the Hispanic GSN is divided into a greater number of communities (N_c) when compared with the white GSN, which can indicate a distinct structural disparity in the ethnoracial pattern in the flow of organs, which needs to be addressed.

Table 1. Network measures of ethnoracial geographic social network focused on recipients.

GSN[a]	N_n [b]	M [c]	M/N_n [d]	CC [e]	L [f]	N_c [g]
All	31,793	266,812	17	0.068	3.968	9
Hispanic	12,025	31,232	5	0.092	5.166	11
Black	16,925	53,697	6	0.126	4.738	12
White	29,606	172,506	12	0.044	4.284	6

[a]GSN: geographical social network.

[b]N_n: number of nodes.

[c]M: links.

[d]M/N_n: average degree.

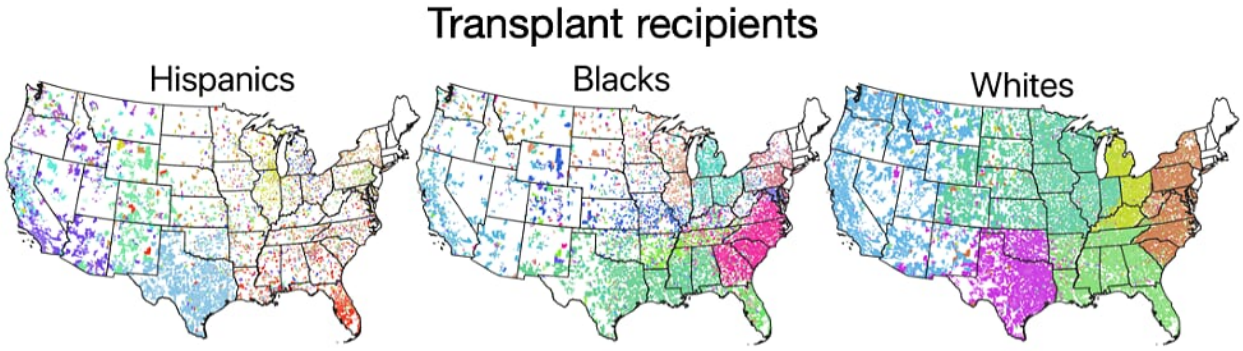
[e]CC: clustering coefficient.

[f]L: average path length.

[g]N_c: number of communities.

By examining the geographic spread of these communities across the United States (Figure 2), one can see that Hispanic communities appear more geographically spread. Although similar white communities are located close to one another and thus form well-defined geographic boundaries, Hispanic communities are more geographically dispersed such that same communities have a higher chance of being located far from each other. This higher geographic spread of communities in the Hispanic GSN along with its higher average path length quantitatively describes unintended differences in the organ allocation mechanism for Hispanic recipients. In principle, organ allocation should be as local as possible according to UNOS.

Figure 2. Ethnic/racial communities of geographic social network (GSN). The communities are extracted from separately generated GSNs from transplant recipients that are Hispanics (left), blacks (center), and whites (right). Minority populations (ie, Hispanics and blacks) experience a greater number of disorganized communities within the United States.



Evaluation of a Sensor of Organ Donation Using Twitter

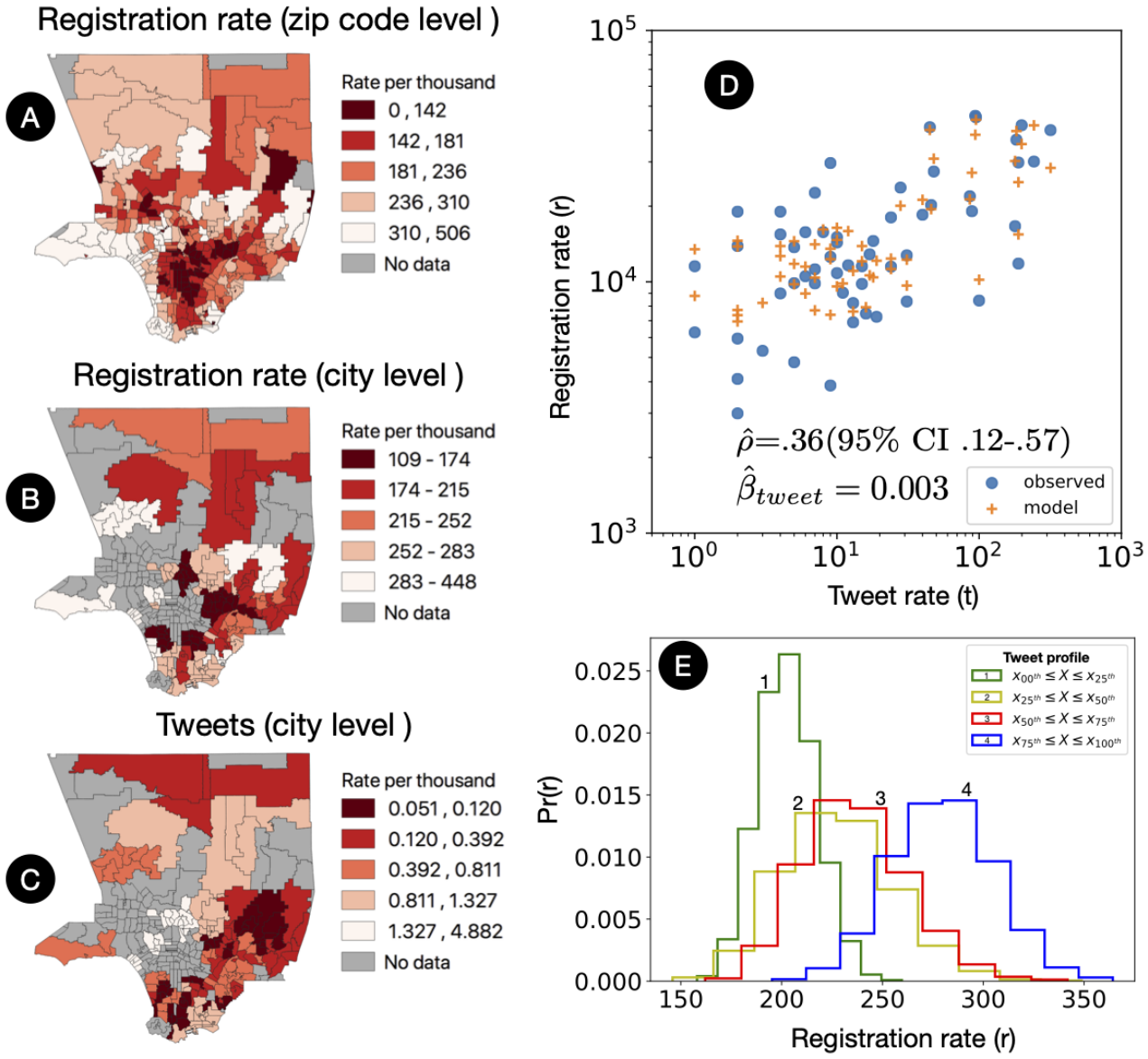
The descriptive statistics of our collected tweets are described in Table 2. Tweets were collected using our real-time organ donation sensor, and organ donation registrations were obtained from the Department of Motor Vehicles in the greater Los Angeles Area (LA County). Our organ donation sensor shows that the number of organ-related tweets are associated with the number of organ donation registrations (Figure 3). After normalizing for the population size, the number of organ donor registrations (Figure 3) are significantly correlated with the number of organ-related tweets at the city level (Figure 3). A Poisson regression predicts that each 10 additional organ-related tweets are associated with a 3.20% (33,933/1,060,403) increase in the number of donor registrations (Figure 3). Similarly, the data-intensive bootstrapping predicts that, on average, the number of organ donor registrations can vary (Figure 3) from 202 (95% CI 176-32) for cities with organ-related tweet rates between 0 and 25 percentiles to 279 (95% CI 231-329) for cities with organ-related tweet rates between 75 and 100 percentiles.

Yearly state-level organ registration data obtained from publicly available Donate Life annual reports from 2009 to 2016 [2] were additionally used to validate that the organ-related tweets collected in 2016 and further aggregated at the state level increasingly correlate with more recent registration data. For instance, organ-related tweets are more correlated with 2016 registrations (r=.81; P<.01) than with 2009 registrations (r=.51; P<.01) and 2012 registrations (r=.70; P<.01).

Table 2. Descriptive statistics of tweets collected by the organ donation Twitter sensor.

Statistic	Value
Data collection start date	April 22, 2015
Data collection end date	May 11, 2016
Data collection number of days	385
Number of collected tweets	134,986
Number of Twitter users	71,947
Average number of tweets per day	350
Average number of tweets per user	1.88
Number of organs mentioned per tweet	1.03
Number of organs mentioned per user	1.13

Figure 3. Association between organ-related tweets and organ donation registrations. (A) Organ donation registrations at the zip code level. (B) Organ donation registrations aggregated at the city level. (C) Organ-related tweets at the city level. (D) Poisson model of donation registration predicted by organ-related tweets after controlling for population size. (E) The profile of organ-related tweet percentile of a city is associated with the organ donation registrations of that city.



Exposure to a Focused Audience

The SNI reached more than 1 million individual users on Facebook (Table 3). Users in social media, including Facebook, can be overrepresented or underrepresented when compared with the actual population. As the targeted audience is increasingly narrowed, such deviation can be intensified. The advertising platform on Facebook provides insights on the targeted audience according to multiple criteria, including gender and socioeconomics (Table 3). For instance, the audience targeted by our SNI had moderately lower household income. However, more women (939,666/1,174,583; 80.00%) were unexpectedly reached than men (234,917/1,174,583; 20.00%).

The educational content associated with the highest clicks per impression (*C/I*) during the first phase of the intervention is defined as the most appealing content. Such content is subsequently used to optimize the intervention in a second phase. This optimization played a key role in exposing the most appealing content to the targeted audience while promoting higher engagement rates per impression.

Table 3. Population demographics targeted by the social network intervention. Overall, the audience targeted by our SNI had moderately lower household income, and more women were reached than men.

Demographic characteristics	Values (n=1,174,583)
Gender, n (%)	
Women	939,666 (80.00)
Men	234,917 (20.00)
Age (women), n (%)	
18-24	58,729 (5.00)
25-34	293,646 (25.00)
35-44	293,646 (25.00)
45-54	234,917 (20.00)
55-64	176,187 (15.00)
>65	58,729 (5.00)
Age (men), n (%)	
18-24	0 (0.00)
25-34	411,104 (35.00)
35-44	411,104 (35.00)
45-54	411,104 (35.00)
55-64	0 (0.00)
>65	0 (0.00)
Household income (USD), n (%)	
30-40	117,458 (10.00)
40-50	176,187 (15.00)
50-75	411,104 (35.00)
75-100	176,187 (15.00)
100-125	117,458 (10.00)
125-150	117,458 (10.00)
150-250	117,458 (10.00)
250-350	0 (0.00)
350-500	0 (0.00)
>500	0 (0.00)
Household ownership, n (%)	
Renter	352,375 (30.00)
Owner	822,208 (70.00)

Efficacy of Exposure and Engagement

Facebook's advertisement platform provided the number of impressions, clicks, and page views daily (Table 4; Figure 4). These measurements are highly correlated, and this high correlation structure increased after optimization (Figure 4). To control for differences between resource utilization after the optimization as measured by the number of impressions, the number of clicks (*C/I*) and page views (*V/I*) were normalized by the number of impressions (Figure 4). Although *C/I* and *V/I* are negatively correlated with before the optimization, both ratios become more positively correlated after the optimization (Figure 4).

The results of the OLS regression indicate that the use of optimization can increase *C/I* (beta=.0213; *P*<.004). For instance, 21,000 clicks can be additionally fostered when exposing 1 million individuals (Table 5 and Figure 4). According to the regression, an additional 21 (95% C 8-35) clicks can be obtained per thousand of impressions after the optimization, with the number of clicks per thousand impressions increasing from 42 (95% CI 35-48) to 63 (95% CI 50-77). One can see a saturation between clicks and impressions. The *C/I* began to saturate as *I* increased, but this saturation was lower after the optimization. Before the optimization, as *I* increased, *C/I* decreased from 41 (95% CI 40-41) to 21 (95% CI 10-31). This saturation vanished after the optimization, and *C/I* has not statistically changed as *I* increased. Conversely, *V/I* was not significantly changed after the optimization. All data can be made available for future studies upon request.

Table 4. Social network intervention before and after optimization. The number of impressions, clicks, and page views provided daily by Facebook's advertisement platform.

Date/period	*I* [a]	*C* [b]	*V* [c]	*C*/*I* [d] (%)	*V*/*I* [e] (%)
All intervention					
Total period	1,174,583	53,988	19,901	4.60	1.69
Pre optimization					
Total period	372,524	10,077	3705	2.71	0.99
August 4	4639	198	102	4.27	2.20
August 5	8831	346	200	3.92	2.26
August 6	11,058	412	204	3.73	1.84
August 7	14,731	544	290	3.69	1.97
August 8	24,697	699	272	2.83	1.10
August 9	28,165	563	237	2.00	0.84
August 10	31,336	778	242	2.48	0.77
August 11	23,904	602	172	2.52	0.72
August 12	21,661	578	172	2.67	0.79
August 13	17,584	501	167	2.85	0.95
August 14	16,884	417	124	2.47	0.73
August 15	22,518	585	198	2.60	0.88
August 16	20,854	523	188	2.51	0.90
August 17	19,964	458	168	2.29	0.84
August 18	18,252	435	161	2.38	0.88
August 19	15,264	353	126	2.31	0.83
August 20	16,552	381	168	2.30	1.02
August 21	17,594	392	148	2.23	0.84
August 22	15,061	528	138	3.51	0.92
August 23	22,975	784	228	3.41	0.99
Post optimization					
Subtotal	802,059	43,911	16,196	5.47	2.02
August 24	53,280	2708	825	5.08	1.55
August 25	54,076	3154	1007	5.83	1.86
August 26	47,259	2778	819	5.88	1.73
August 27	55,165	3067	898	5.56	1.63
August 28	67,832	3882	1485	5.72	2.19
August 29	72,089	4243	1664	5.89	2.31
August 30	79,789	4679	1721	5.86	2.16
August 31	88,074	4967	2041	5.64	2.32
September 1	96,850	5118	2118	5.28	2.19
September 2	88,455	4770	1975	5.39	2.23
September 3	99,190	4545	1643	4.58	1.66

[a]*I*: number of impressions.

[b]*C*: clicks.

[c]*V*: page views.

[d]*C/I*: clicks per impression.

[e]*V/I*: page views per impression.

Figure 4. Effectiveness of the social network intervention. (A-B) The daily metrics of the impressions, clicks, page views, as well as their normalized versions, clicks per impression, and page views per impression. (C) The rate of clicks per impression and page views per impression became more positively associated after the optimization. (D-E) The regression analysis implicates the use of optimization plays a key role in positively affecting clicks per impression and page views per impression. For instance, after the optimization, clicks per impression was 0.0213 higher.

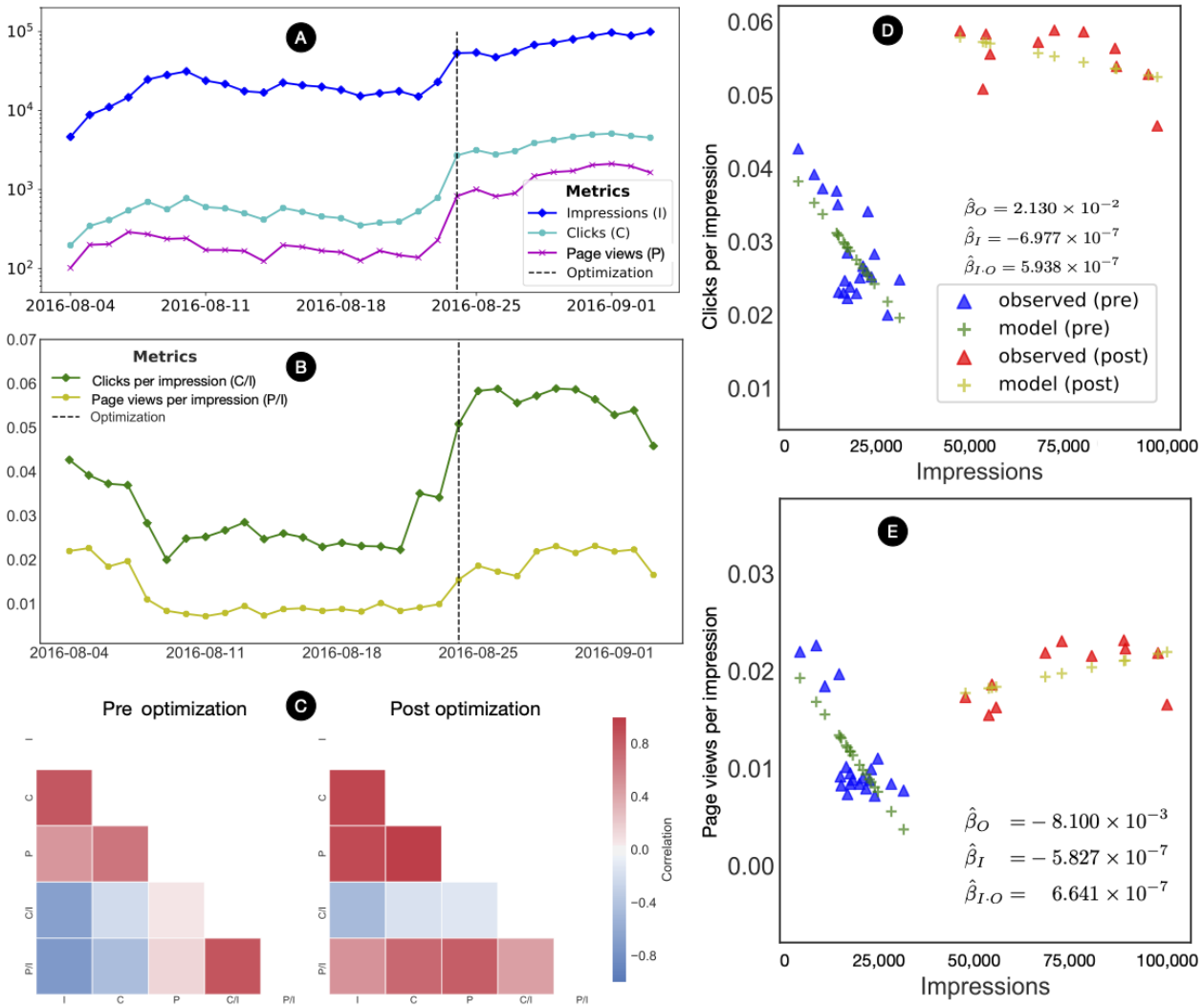


Table 5. Results of the ordinary least squares regression of clicks per impression and page views per impression relative to the number of impressions and optimization.

Estimator	Coefficient	SE	*P* value
Clicks per impression (*C/I*)			
Constant	0.0415	0.003	<.001
Optimization (*O*)	0.0213	0.007	.004
Impressions (*I*)	−6.977e-07	<0.001	<.001
Optimization×impressions (*O*I*)	5.938e-07	<0.001	.003
F statistic (*df*)	85.29 (3,27)	—[a]	<.001
R^2	0.905	—	—
Adjusted R^2	0.894	—	—
Page views per impression (*V/I*)			
Constant	0.0220	0.002	<.001
Optimization (*O*)	−0.0081	0.005	.10
Impressions (*I*)	−5.827e-07	<0.001	<.001
Optimization×impressions (*O*I*)	<0.001	<0.001	<.001
F statistic (*df*)	25.45 (3,27)	—	<.001
R^2	0.739	—	—
Adjusted R^2	0.710	—	—

[a]Not applicable.

Discussion

Principal Findings

In this study, we proposed a framework for a large-scale community-based intervention using social media: SNI. Our framework demonstrated an affordable and effective application of social media in rapidly exposing and engaging large populations to address the disproportionate lack of awareness regarding organ donation among minorities. In a period of 1 month, our SNI was able to engage 1 million individuals, which is a much larger audience compared with traditional community-based interventions focused on health education through more costly and rigid frameworks. These traditional interventions relied heavily on health professional interactions with communities to disseminate generalized information without taking into account specific community information such as demographics, optimally relatable material, or highly shareable content through established social networks. A larger audience in conjunction with tailored content provides an ideal platform to effectively engage a target population while potentiating a shift toward positive attitudes regarding organ donation.

By implicating clicks as a form of positive attitude and engagement with organ donation, we showed that targeting a focused audience with tailored content is key to making an intervention more effective. The higher the number of clicks per impression on certain Web-based materials implied that some content had greater impacts on the target audience in motivating engagement with the material. The most effective content presented to the target audience was automatically learned during the intervention and determined to be an optimization priority. Precisely, 21,000 additional clicks were obtained because of the optimization alone, which shows the efficacy and power of an optimizable data-driven network.

A network-based intervention approach has shown the ability to increase target audience engagement with organ donation compared with traditional community-based approaches. This directly potentiates an increase in the proportion of target audience donors at a particular location. The broader impact of this form of intervention results in network changes that can bolster an established organ donor community with every additional organ donor, leading to a higher clustering coefficient and a decrease in the average path length for organ transplantation within a particular GSN.

Limitations

The major limitation of our current SNI is its inability to measure the actual donor registrations that were obtained as a direct result of the intervention. Our SNI focused on the efficacy of eliciting a simple behavioral action as a proxy for a shift toward positive attitudes regarding organ donation, namely, a click on the organ donor registration site link.

Another limitation is that the data collected in this study are not recent: the organ donation data from our Twitter sensor were collected from April 2015 to May 2016, and the intervention data from Facebook were collected from August 2016 to September 2016. In the study of organ donation, timely access to longitudinal and high-resolution data on organ donation registrations is a major challenge. Additionally, we only had access to yearly state-level organ registration data obtained from

publicly available Donate Life annual reports from 2009 to 2016 [2]. However, we have demonstrated in our results that organ-related tweets are correlated with registration rates at the city level even after controlling for population and additionally validated that the organ-related tweets collected in 2016 increasingly correlate with more recent registration data.

Our results were limited to Hispanics and may not necessarily generalize to other minority populations, such as Asians and American Indians. Future studies will be directed to each specific population with their respective community-driven study designs.

Conclusions

Organ transplantation remains the only life-saving therapy option for patients with end-stage organ failure. However, the lack of organ donors limits the availability of organs for transplant. Although the numbers of organ donors and transplantations in the United States have doubled over the past 20 years, the demand for organs continues to exceed the supply. In 2016, there were over 30,000 solid organ transplantations; however, more than 120,000 people remain on waiting lists for transplants. Associated health care costs related to the management of end-stage disease and associated disabilities outstrip those of transplantations. Therefore, an increase in awareness is needed particularly among minority populations.

At the center of our intervention is the recognition that sociocultural dynamics greatly affect what people incorporate into their own beliefs. Prior campaigns that successfully addressed minority-related organ donation disparity relied on grassroots initiatives and interventions that addressed social and psychological influences of an inadequate knowledge base, misinformation, and medical distrust [17]. We built upon this community-oriented design by expanding an individual's social network to incorporate their social media circles. Sociocultural influences and personal experiences have been found to drive engagement with the issue of organ donation during prior grassroots campaigns targeting the African-American minority demographic [17-19]. Taking this into account, we tailored our intervention content to appeal to the target minority population on an intimate level by utilizing personal accounts and relatable statistics while providing the targeted audience with the tools to propagate their newly acquired information within their social context [17].

In this work, we proposed a framework for SNI that is both tailored and large-scale using social media. First, we identified structural disparities in organ transplantation among minority groups using a network-based analysis. Next, we created a digital sensor to monitor population awareness about organ donation using social media and validated the sensor using donation registration data. Afterward, we created an intervention campaign to target a focused audience with educational contents regarding organ donation. Finally, we optimized our SNI to target the contents that were automatically identified as more tailored to our focused audience. Therefore, we proposed a conceptual framework (Figure 1) that puts all these separate pieces together to enable a more systemic approach to effective health literacy interventions.

It is important to note that the network analysis and community detection may be more appropriate for a system-wide evaluation of the UNOS allocation of organs. Instead of measuring individual components of the system such as the proportion of donors at specific locations, network-based analysis can give us systems-level measures such as the average path length. Increasing proportions of donors at individual, possibly disconnected, locations might not necessarily improve the average path length at the system level.

We have shown that social media can be used as a sensor for organ donation awareness. Such a sensor has the potential to monitor organ donation awareness in real time at large-scale. In addition, social media can serve as a platform for delivering large-scale, community-based interventions to raise awareness while improving public attitudes and concern for a public health issue such as organ donation. For future studies, we aim to design a longer SNI capable of capturing changes in organ donation awareness on Twitter because of interventions on Facebook. Likely, these changes will also be associated with organ donation registrations.

Acknowledgments

This work was partially funded by the Rosenfeld Heart Foundation. The sponsor had no role in the study. The authors would like to thank Jung Min Choi for his insights into social theories utilized in their intervention and Thomas Mone from OneLegacy for his advice about organ donation interventions.

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 01;150(3):252-259. [doi: 10.1001/jamasurg.2014.2038] [Medline: 25629390]
2. Donate Life America. 2017 Annual Report. URL: https://www.donatelife.net/wp-content/uploads/2016/06/2017_AnnualUpdate_singlepages_small.pdf [accessed 2019-02-19] [WebCite Cache ID 76JSQwO37]
3. Gordon EJ, Lee J, Kang RH, Caicedo JC, Holl JL, Ladner DP, et al. A complex culturally targeted intervention to reduce Hispanic disparities in living kidney donor transplantation: an effectiveness-implementation hybrid study protocol. BMC Health Serv Res 2018 Dec 16;18(1):368 [FREE Full text] [doi: 10.1186/s12913-018-3151-5] [Medline: 29769080]

4. Thomas SS, D'Alessandro DA. Traumatic brains and broken hearts: mending the donor shortage in cardiac transplantation. J Am Coll Cardiol 2017 Sep 05;70(10):1259-1261 [FREE Full text] [doi: 10.1016/j.jacc.2017.07.756] [Medline: 28859789]
5. Watson CJE, Dark JH. Organ transplantation: historical perspective and current practice. Br J Anaesth 2012 Jan;108(Suppl 1):i29-i42 [FREE Full text] [doi: 10.1093/bja/aer384] [Medline: 22194428]
6. Dierselhuis M, Goulmy E. The relevance of minor histocompatibility antigens in solid organ transplantation. Curr Opin Organ Transplant 2009 Aug;14(4):419-425. [doi: 10.1097/MOT.0b013e32832d399c] [Medline: 19444105]
7. Kransdorf EP, Pando MJ, Gragert L, Kaplan B. HLA population genetics in solid organ transplantation. Transplantation 2017 Sep;101(9):1971-1976. [doi: 10.1097/TP.0000000000001830] [Medline: 28832450]
8. Melanson TA, Hockenberry JM, Plantinga L, Basu M, Pastan S, Mohan S, et al. New Kidney Allocation System Associated With Increased Rates Of Transplants Among Black And Hispanic Patients. Health Aff (Millwood) 2017 Dec 01;36(6):1078-1085 [FREE Full text] [doi: 10.1377/hlthaff.2016.1625] [Medline: 28583967]
9. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000;15(3):267. [doi: 10.1093/heapro/15.3.259]
10. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15(Suppl 2):9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
11. Hälleberg Nyman M, Nilsson U, Dahlberg K, Jaensson M. Association between functional health literacy and postoperative recovery, health care contacts, and health-related quality of life among patients undergoing day surgery: secondary analysis of a randomized clinical trial. JAMA Surg 2018 Aug 01;153(8):738-745 [FREE Full text] [doi: 10.1001/jamasurg.2018.0672] [Medline: 29710226]
12. Sewell AA. Inequality: live poor, die young. Nature 2017 May 18;545(7654):286-287. [doi: 10.1038/545286a]
13. Gordon EJ, Shand J, Black A. Google analytics of a pilot mass and social media campaign targeting Hispanics about living kidney donation. Internet Interv 2016 Nov;6:40-49 [FREE Full text] [doi: 10.1016/j.invent.2016.09.002] [Medline: 30135813]
14. Jacob Arriola KR, Redmond N, Williamson DH, Thompson NJ, Perryman JP, Patzer RE, et al. A community-based study of giving ACTS: organ donation education for African American adults. J Natl Med Assoc 2019 Apr;111(2):185-192. [doi: 10.1016/j.jnma.2018.09.005] [Medline: 30327141]
15. Miller DB, Cage JL, Nowacki AS, Jackson B, Modlin CS. Health literacy (HL) & health-related quality of life (HRQL) among minority men. J Natl Med Assoc 2018 Apr;110(2):124-129. [doi: 10.1016/j.jnma.2017.10.001] [Medline: 29580445]
16. Smith SG, Jackson SE, Kobayashi LC, Steptoe A. Social isolation, health literacy, and mortality risk: findings from the English longitudinal study of ageing. Health Psychol 2018 Feb;37(2):160-169 [FREE Full text] [doi: 10.1037/hea0000541] [Medline: 29172607]
17. Irving MJ, Tong A, Jan S, Cass A, Rose J, Chadban S, et al. Factors that influence the decision to be an organ donor: a systematic review of the qualitative literature. Nephrol Dial Transplant 2012 Jun 21;27(6):2526-2533. [doi: 10.1093/ndt/gfr683] [Medline: 22193049]
18. Callender CO, Miles PV. Minority organ donation: the power of an educated community. J Am Coll Surg 2010 May;210(5):708-15, 715 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2010.02.037] [Medline: 20421035]
19. Callender CO, Koizumi N, Miles PV, Melancon JK. Organ Donation in the United States: The Tale of the African-American Journey of Moving From the Bottom to the Top. In: Transplant Proc. 2016 Sep Presented at: Transplant Proc ;48(7); 2016; Boston, MA p. 2392-2395. [doi: 10.1016/j.transproceed.2016.02.094]
20. Sadasivam RS, Volz EM, Kinney RL, Rao SR, Houston TK. Share2Quit: web-based peer-driven referrals for smoking cessation. JMIR Res Protoc 2013 Sep 24;2(2):e37 [FREE Full text] [doi: 10.2196/resprot.2786] [Medline: 24067329]
21. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
22. Centola D. The spread of behavior in an online social network experiment. Science 2010 Sep 03;329(5996):1194-1197 [FREE Full text] [doi: 10.1126/science.1185231] [Medline: 20813952]
23. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 02;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
24. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: a randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
25. McLean SA, Wertheim EH, Masters J, Paxton SJ. A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. Int J Eat Disord 2017 Jul 28;50(7):847-851. [doi: 10.1002/eat.22708] [Medline: 28370321]
26. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
27. Proestakis A, di Sorrentino EP, Brown HE, van Sluijs E, Mani A, Caldeira S, et al. Network interventions for changing physical activity behaviour in preadolescents. Nat Hum Behav 2018 Oct 24;2(10):778-787 [FREE Full text] [doi: 10.1038/s41562-018-0436-y] [Medline: 30886902]
28. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]

29. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions-evidence from Twitter. J Am Med Inform Assoc 2017 May 01;24(3):496-502. [doi: 10.1093/jamia/ocw133] [Medline: 27707822]
30. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Social Science & Medicine 2000 Sep;51(6):843-857. [doi: 10.1016/S0277-9536(00)00065-4]
31. Kim DA, Hwong AR, Stafford D, Hughes DA, O'Malley AJ, Fowler JH, et al. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. Lancet 2015 Jul;386(9989):145-153. [doi: 10.1016/S0140-6736(15)60095-2]
32. Lyson HC, Le GM, Zhang J, Rivadeneira N, Lyles C, Radcliffe K, et al. Social media as a tool to promote health awareness: results from an online cervical cancer prevention study. J Cancer Educ 2019 Aug;34(4):819-822. [doi: 10.1007/s13187-018-1379-8] [Medline: 29948924]
33. Venugopal S, Stoner E, Cadeiras M, Menezes R. Understanding organ transplantation in the USA using geographical social networks. Soc Netw Anal Min 2012 Dec 7;3(3):457-473. [doi: 10.1007/s13278-012-0089-1]
34. Barabási AL, Pósfai M. Network Science. Cambridge, United Kingdom: Cambridge University Press; 2016.
35. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
36. Pacheco D, Pinheiro D, Cadeiras M, Menezes R. Characterizing Organ Donation Awareness from Social Media. 2017 Apr Presented at: 33rd IEEE International Conference on Data Engineering (ICDE); 2017; San Diego, CA p. 19-22. [doi: 10.1109/icde.2017.225]
37. Verdone M. Python Package Index. 2019. The Minimalist Twitter API for Python. URL: https://pypi.org/project/twitter/
38. Python Package Index. Geopy. URL: https://pypi.org/project/geopy/
39. Reibel M, Olmo C, Andrada S, Koertzen J. Deep demographics: understanding local variation in donor registration. Prog Transplant 2016 Jun 04;26(2):191-198. [doi: 10.1177/1526924816640670] [Medline: 27207409]
40. Donate California. Registration. URL: https://register.donatelifecalifornia.org/register/
41. Bucher T. Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media Soc 2012 Apr 08;14(7):1164-1180. [doi: 10.1177/1461444812440159]
42. Bhagat S, Burke M, Diuk C, Filiz I, Edunov S. Facebook. 2016. Three and a half degrees of separation. URL: https://research.fb.com/three-and-a-half-degrees-of-separation/ [accessed 2019-03-04] [WebCite Cache ID 76d4tCu6q]

Abbreviations

API: application programming interface
GSN: geographical social network
OPO: organ procurement organization
OLS: ordinary least squares
SNI: social network intervention
UNOS: United Network for Organ Sharing

Edited by G Eysenbach; submitted 16.05.19; peer-reviewed by J Brixey, C Callendar; comments to author 30.08.19; revised version received 24.10.19; accepted 12.11.19; published 14.01.20.

Please cite as:
Murphy MD, Pinheiro D, Iyengar R, Lim G, Menezes R, Cadeiras M
A Data-Driven Social Network Intervention for Improving Organ Donation Awareness Among Minorities: Analysis and Optimization of a Cross-Sectional Study
J Med Internet Res 2020;22(1):e14605
URL: https://www.jmir.org/2020/1/e14605
doi:10.2196/14605
PMID:31934867

Review

Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review

Franziska Burger[1], MSc; Mark A Neerincx[1,2], PhD; Willem-Paul Brinkman[1], PhD

[1]Delft University of Technology, Delft, Netherlands

[2]Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO), Department of Perceptual and Cognitive Systems, Soesterberg, Netherlands

Corresponding Author:
Franziska Burger, MSc
Delft University of Technology
van Mourik Broekmanweg 6
Delft, 2628XE
Netherlands
Phone: 31 0645340352
Email: f.v.burger@tudelft.nl

Abstract

Background: Electronic mental (e-mental) health care for depression aims to overcome barriers to and limitations of face-to-face treatment. Owing to the high and growing demand for mental health care, a large number of such information and communication technology systems have been developed in recent years. Consequently, a diverse system landscape formed.

Objective: This literature review aims to give an overview of this landscape of e-mental health systems for the prevention and treatment of major depressive disorder, focusing on three main research questions: (1) What types of systems exist? (2) How technologically advanced are these systems? (3) How has the system landscape evolved between 2000 and 2017?

Methods: Publications eligible for inclusion described e-mental health software for the prevention or treatment of major depressive disorder. Additionally, the software had to have been evaluated with end users and developed since 2000. After screening, 270 records remained for inclusion. We constructed a taxonomy concerning software systems, their functions, how technologized these were in their realization, and how systems were evaluated, and then, we extracted this information from the included records. We define here as functions any component of the system that delivers either treatment or adherence support to the user. For this coding process, an elaborate classification hierarchy for functions was developed yielding a total of 133 systems with 2163 functions. The systems and their functions were analyzed quantitatively, with a focus on technological realization.

Results: There are various types of systems. However, most are delivered on the World Wide Web (76%), and most implement cognitive behavioral therapy techniques (85%). In terms of content, systems contain twice as many treatment functions as adherence support functions, on average. Furthermore, autonomous systems, those not including human guidance, are equally as technologized and have one-third less functions than guided ones. Therefore, lack of guidance is neither compensated with additional functions nor compensated by technologizing functions to a greater degree. Although several high-tech solutions could be found, the average system falls between a purely informational system and one that allows for data entry but without automatically processing these data. Moreover, no clear increase in the technological capabilities of systems showed in the field, between 2000 and 2017, despite a marked growth in system quantity. Finally, more sophisticated systems were evaluated less often in comparative trials than less sophisticated ones (OR 0.59).

Conclusions: The findings indicate that when developers create systems, there is a greater focus on implementing therapeutic treatment than adherence support. Although the field is very active, as evidenced by the growing number of systems developed per year, the technological possibilities explored are limited. In addition to allowing developers to compare their system with others, we anticipate that this review will help researchers identify opportunities in the field.

(J Med Internet Res 2020;22(1):e12599) doi:10.2196/12599

KEYWORDS

eHealth; major depressive disorder; technology; systematic review

Introduction

Between 2000 and 2017, researchers have reported more than 100 software interventions for depression in the scientific literature. Although all these systems have the same objective, they vary widely in both content and in the way the content is delivered. Taken together, they thus form a diverse landscape. But what does this landscape actually look like? The purpose of this literature review is to map the terrain by exploring the technological state of the art of electronic mental (e-mental) health interventions for depression.

The systems under study here strive to meet a globally growing need for depression care. The illness affects approximately 300 million people worldwide [1]. Its high lifetime prevalence and high disease burden are further exacerbated by additional episodes often following the first. This renders the pervasive provision of treatment and prevention means imperative. However, the World Health Organization estimates that, currently, half of those suffering from depression are receiving inadequate or no treatment [1].

Information and communication technology (ICT) may present a viable solution to the shortage. The rapid dissemination of ICT over the course of the past two decades has led researchers to explore the provision of therapeutic content on these platforms. Unlike face-to-face treatment, such support systems are scalable, easily accessible, cheap, and standardized, and they can reduce the fear of stigmatization, as they can be used in private and at one's own convenience [2]. In addition to these benefits, numerous meta-analyses attest to the effectiveness of the interventions [3-5].

As a consequence of the high research interest, many systems have been developed to treat or prevent depression. Each system presents a unique solution. In light of this, several recent literature surveys point out that an analysis of the system landscape is in order, as there is little insight into the makeup of systems [2,6,7]. Where systems have been reviewed to date, authors have typically adopted one of two core perspectives. Syntheses with a *clinical psychology* perspective have addressed the effectiveness of different types of interventions [3,8,9]. Syntheses with a *(persuasive) technology* perspective, on the other hand, have addressed the functionality of systems, such as persuasive technology elements [7] or communication modality [10]. This systematic literature review takes the latter perspective. However, rather than studying in depth the implementation or impact of a specific type of function, it compares entire systems on their technological implementation. In doing so, e-mental health systems for depression are regarded as compositions of functions and assessed in terms of their technological realization. The support systems reported in the literature thus form the population under study. The main goal of this review is then to provide a comprehensive overview of the system landscape and its technological state. In addition, it identifies some of the challenges and opportunities for the field. However, linking the degree to which systems present high-tech solutions with clinical outcomes is outside of the scope of this review. Nevertheless, with the introduced system characterization and technological sophistication metric, a first step toward such studies is taken. From the extensive, domain-specific analysis presented here, we particularly expect researchers setting out to develop or study support systems for depression to benefit. It allows them to compare their system with those already in use and to identify underexplored aspects of these systems. To this end, the following three research questions are addressed:

1. What types of ICT systems for the treatment and prevention of depression have been developed?
2. How technologized are these systems?
3. How has the system landscape evolved between 2000 and 2017?

Methods

Literature Identification and Coding

In this section, we focus on the literature search and filtering as well as coding of data pertaining to the analyses in this study. A detailed account of the construction, the structure, and the information contained in the open-access, relational database that was created for this analysis can be found in the documents [11,12].

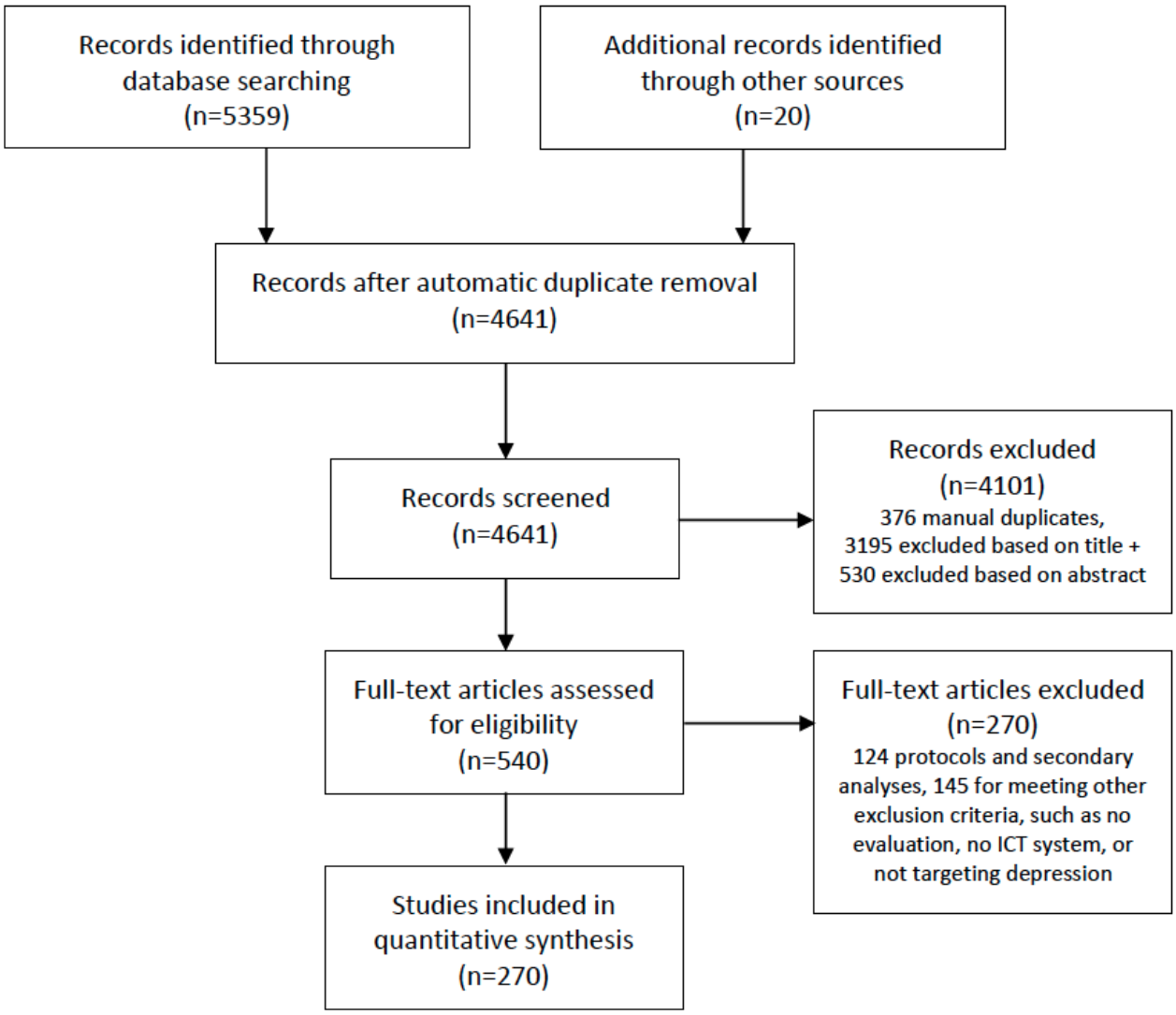
Identification

The exhaustive search for potentially relevant literature made use of 3 databases: Scopus, PubMed, and Web of Science. It included English language journal articles, conference papers, and theses published between 2000 and 2017, presenting primary research that was conducted with support systems for the prevention or treatment of major depressive disorder or dysthymia in adults. To ensure that systems were actually created and functional at some point, we only considered the literature that reported the results of a system evaluation with end users. Therefore, systems that only had published study protocols available at the time of the search (early 2017) did not qualify. Lists of search terms comprised words around the following concepts that were central to the research interest: *ICT, Health Condition, Purpose, Evaluation* (Multimedia Appendix 1). They were expanded with controlled vocabulary terms, where applicable. Systems met exclusion criteria if they were (1) employing technology for mediated communication, (2) targeting children, postpartum or pregnant women, caregivers of depressed patients, or patients with comorbid psychotic conditions, (3) only aiming to reduce stigma, (4) serving only as diagnostic tools or decision aids, (5) addressing only antidepressant treatment, and (6) having an otherwise too narrow scope, for example, a system developed for a single patient with a specific combination of comorbid conditions.

The 3 queried databases returned a total of 5359 documents. Forward and backward reference searches on previous literature reviews and meta-analyses yielded an additional 20 records. After the removal of duplicates, 4256 records remained for

screening. A lenient inclusion protocol at the title and abstract stages allowed for the inclusion of as many articles as possible concerning a system. Therefore, the exclusion of articles describing study protocols and secondary analyses only occurred at the full paper screening, but they were kept as additional references for clarification purposes. The first author, with a cognitive science background, screened all records at the title, abstract, and full-text stages (see PRISMA [13] diagram in Figure 1). A second, independent coder, with a computer science background, double coded a random selection at each stage. Intercoder agreement ranged from 80% to 84%, with moderate-to-substantial intercoder reliability (Cohen kappa between 0.50 and 0.69). Multimedia Appendix 2 includes a complete list of all 270 articles included in the final synthesis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of the screening process, as completed by the first author. ICT: information and communication technology.



Coding

To provide an overview of the aspects of software systems for depression considered here, a simplified taxonomy is presented in Multimedia Appendix 3. The extraction of information resulted in 45 coded attributes. These were either low-inference attributes, that is, the information could be directly copied from the paper, or high-inference attributes, that is, the coder needed to make inferences to arrive at the information. Second coders were neither used to refine the coding procedure nor to obtain a more reliable dataset. However, second coders did double code samples of the high-inference attributes to assess the reliability of the first coder, and the intercoder reliability measures reported here are to be regarded as an indication thereof. All analyses are based on the coding of the first author only.

A key task in the coding process was the division of systems into elementary functional parts, that is, functions. Herein, the focus was limited to functions pertaining to the higher-level layers of software architecture. For example, in the layered software architecture described in the Microsoft Application Architecture Guide [14], the functions would be located in the presentation and application layers. Cross-cutting concerns, such as security, were not considered. Additional criteria by which to evaluate software quality, for example maintainability, integration with other software, or software reliability, are also beyond the scope of this work. The construction of a

classification hierarchy (Figure 2) preceded the coding process. At the fourth and highest level, two types of functions are possible: *intervention* functions, which aim to reduce depressive symptomatology in users, and *support* functions, which aim to increase adherence of the user to the intervention. An example of an intervention function would be the positive psychology exercise to count one's blessings every night, whereas an example of a support function would be to send text message reminders to encourage the user to engage with the system. At the third level, support functions further split into helping the user in (1) planning the intervention, (2) executing the intervention, (3) self-monitoring, or (4) connecting with other supportive people. A total of 2 more refined classification levels follow. At the lowest level, 41 classifications make up the support functions (Multimedia Appendix 4) and 145 classifications make up the intervention functions (Multimedia Appendix 5). Inspiration for the lowest-level support functions came largely from persuasive technology design frameworks [15-18], whereas therapy manuals (eg, [19]) inspired the lowest-level intervention functions. These are often linked to therapeutic intervention frameworks, for example, Activity Planning is a technique of Behavioral Therapy. The intervention frameworks finally cluster into 8 *therapies* (Multimedia Appendix 5).

A second coder with a background in clinical psychology double coded two parts of the function identification task. The first part required spotting functions in the system description. Taking the functions that were found by the first coder as ground truth, interrater reliability was moderate on this part (φ=0.54, with a specificity [20] of d'=2.31). The second part required labeling snippets of text that the first coder had identified as functions. For this part, interrater reliability on the 4 different function classification levels (Figure 2) was good on average (=0.63), ranging from moderate (κ=0.55) to good (κ=0.72).

Another key coding task concerned rating the degree to which each function was technologized. A set of scales, the e-mental Health Degree of Technological Sophistication (eHDTS) rating scales (Multimedia Appendix 6), were developed specifically for this task. They include one scale for intervention functions and four separate scales for each of the four types of support functions. Conceptually, the scales range from *offline* to *responsive on content* (Table 1). Although the emphasis in the interpretation of the eHDTS scales throughout this work is placed on the *interactivity* aspect, the actual scales are broader, also covering aspects such as responsiveness, personalization, data analysis, and data presentation. From here on, when directly describing the technological realization of systems or functions as measured by the scale, we refer to it as *technological sophistication*. In coding, a conservative approach ensured that the lower degree was assigned in case of doubt. Reliability levels were acceptable, with a mean correlation of 0.66 between coders. Furthermore, concurrent validity of the scales was supported by on-average moderate correlations (=0.53) between ratings on these scales and ratings on an unlabeled ordinal scale, that is, leaving it open to coders to decide what the different levels of technological sophistication entail.

Finally, one coder was provided with a list of function descriptions from all function types, without the function type label, and asked to assign a rating of technological sophistication to these on an unlabeled ordinal scale from 0 to 5 (uninformed). After two weeks, he was again invited to code the same functions with the appropriate scale for each function and each scale level defined (informed). The correlation between the uninformed and the informed rating (r=0.47) provided some indication that, although each function type had its own eHDTS scale, the five scales were sufficiently similar to allow for aggregation and cautious comparisons on a system level.

Three low-inference attributes coded were the system version, the system build year, and the evaluation quality. A *version* was defined as a modification of the system offering different functionality. For example, Lemma et al created a version with human support and a version without it [21], whereas Currie et al offer different versions to support female or male patients by providing extra content for women [22]. However, a system with an adaptive user interface based on gender was not regarded as two versions; it was regarded as one with a tailoring function. The system *build year* denotes the year in which systems were finalized, that is, the earliest year of operation mentioned in the earliest publication on the earliest version (versions and systems are simply referred to as *systems* or *software* for legibility from here on. Most analyses to follow were conducted on the body of versions rather than systems. It is made explicit when this is not the case). Finally, the *evaluation quality* received a binominal coding of high and low. A high quality meant that the system was evaluated in a comparative trial, whereas a low quality meant that it was evaluated in a single-group trial. Comparative trials encompassed randomized controlled trials, randomized comparative trials, and nonrandomized comparative trials.

Figure 2. The top three levels of the function classification tree as well as the percent agreement and Cohen's kappa for the function classification task at each of the levels. Level 0 of the tree is specified in Multimedia Appendix 2 (support functions) and Multimedia Appendix 3 (intervention functions). CBT: cognitive behavioral therapy.

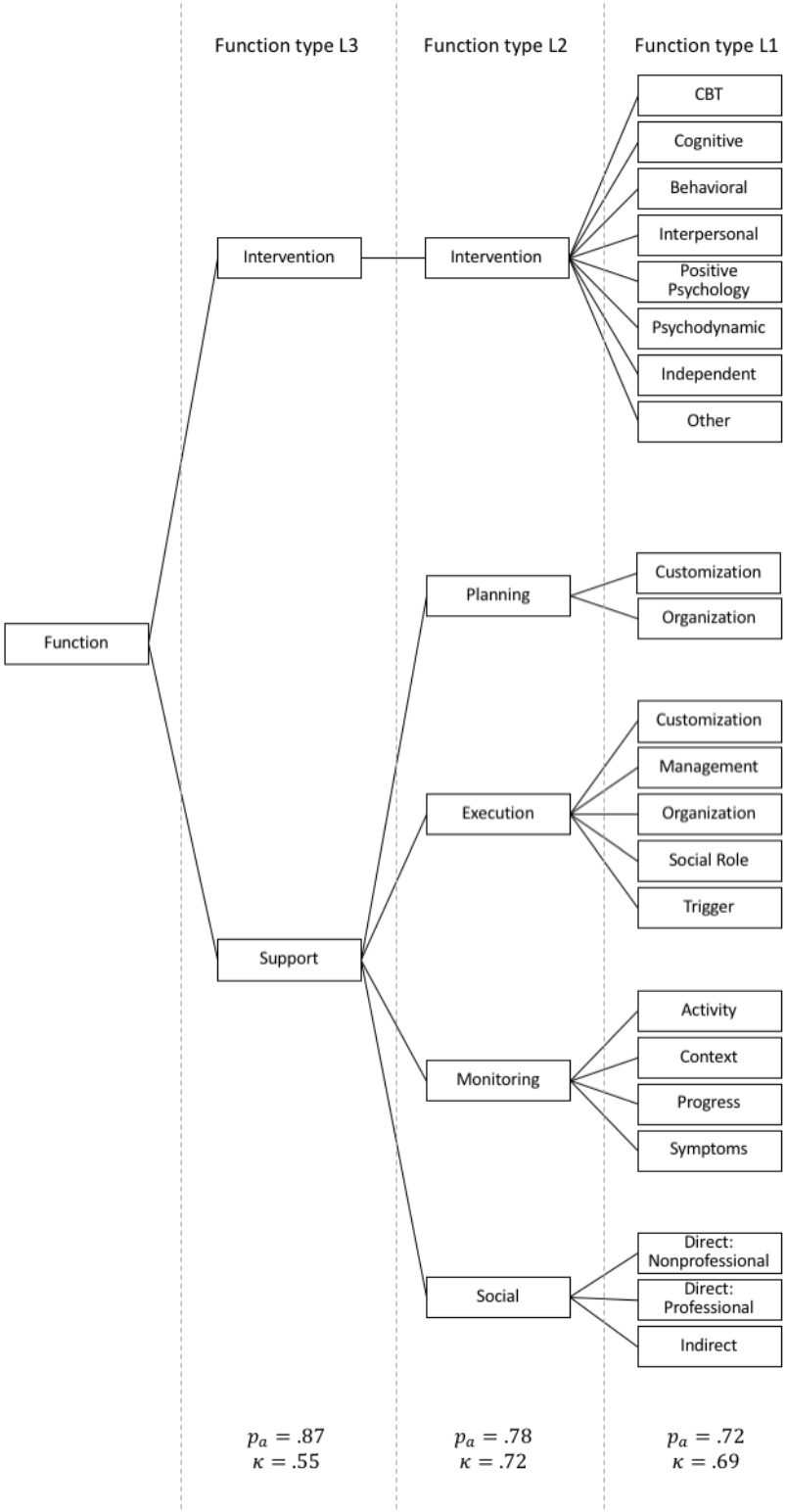


Table 1. The degrees of the e-mental health degree of technological sophistication rating scale, abstracted over the 5 different instantiations of this scale. A diary function serves as a hypothetical example. It should be noted that this is an abstract summary of the levels across several scales. It therefore does not capture the entire technological breadth of the different scales.

Degree	Definition	Example
0: Offline	The function is not provided through the system at all or is fully carried out by a human.	Diary sent by postal mail
1: Informational	The function is provided in an informational manner.	Diary can be downloaded as PDF
2: Data entry	The function is provided in an interactive manner but without processing of input from the user.	Diary can be filled on the Web and saved
3: Form response	The function is provided in an interactive manner with processing of meta-information	Web-based diary that responds to the duration of typing
4: Content response	The function is provided in an interactive manner with processing of the content of user input.	Web-based diary that responds to the sentiment of text, which the user has written, for example, "It appears that this was a very negative experience for you."

Statistical Analysis

We conducted quantitative analyses with R version 3.5. All data and the full analysis script are permanently stored for public access on a national database for research data with the 4TU Center for Research Data in the Netherlands [23]. Where distributions deviated markedly from normality, nonparametric tests were used. Furthermore, we report 2 estimated R^2 effect size measures where R^2 cannot be calculated exactly. For logistic regression models, Nagelkerke pseudo R^2 [24] was chosen, whereas, for multilevel models, the Level1 R^2, as proposed by Snijders and Bosker [25], was computed. When used, these are indicated as Nagelkerke R^2 and Level1 R^2, respectively.

Characterization

To characterize systems, we regarded their composition in terms of functions and how systems differed depending on such factors as guidance or system purpose, that is, prevention or treatment. A Wilcoxon rank sum test compared the number of intervention functions with the number of support functions per system. In addition, two logistic regression models were fit. One determined whether a certain system purpose was more commonly occurring within a certain therapy type. The other tested whether autonomous or guided systems are represented to different degrees depending on purpose. Systems that include human guidance naturally have more functions, as guidance needs to be facilitated by the system somehow. This takes place by way of the direct social support functions. Thus, to allow for a fair comparison of the number of functions in autonomous versus guided systems, direct social support functions were excluded for the following three analyses. First, a linear regression examined the relationship between guidance and the number of functions of a system. Second, two more detailed analyses in the form of Wilcoxon rank sum tests considered this relationship separately for intervention and support functions.

Technological Sophistication

Technological sophistication was compared among the different types of functions, different types of systems, and different evaluation qualities. A correlation assessed the relationship between system size and technological sophistication, and linear regression models gave insight into the link between technological sophistication on the one hand and evaluation quality, guidance, or system purpose on the other. To contrast support and intervention functions, a multilevel linear model was fit using the function type as a fixed effect and allowing for random intercepts per system. Similarly, a 1-way analysis of variance checked for differences in technological sophistication among the four different support types.

Developments Over Time

Changes over time could take place both across and within systems. A total of two linear regression models examined development in size and technological sophistication across systems. Moreover, three multilevel linear models allowed studying development within systems. They determined whether size, technological sophistication, and evaluation quality changed across versions. Random intercepts modeled the nested relationship of versions within systems.

Results

Characterization

In total, 133 systems with 259 versions were identified. Coding these systems on their key attributes led to the characterization presented in Table 2.

Versions

Systems had 2 versions on average, but more than two-thirds (69.2%, 92/133) only had 1 version. Thus, most systems seem to have been developed for a single research project. Only 10 systems had 5 or more versions, for example, The Sadness Program with 13 versions, MoodGYM with 15 versions, and the Well-being Course with 18 versions.

Table 2. The distributions over technology-related key attributes of depression support system versions.

Technology	Value
Number of versions[a] (N=133), mean (SD)	2.0 (2.5)
Technology (N=259), n (%)	
Offline	69 (26.6)
World Wide Web	196 (75.7)
Email	112 (43.2)
Telephone	53 (20.5)
Computer	28 (10.8)
Text message	17 (6.6)
Mobile	16 (6.2)
App	14 (5.4)
Sensors	7 (2.7)
Social media	6 (2.3)
Virtual agent	5 (1.9)
Interactive voice response	5 (1.9)
CD/DVD	5 (1.9)
Virtual reality	2 (0.8)
Undefined	4 (1.5)
Support type (N=259), n (%)	
Autonomous	123 (47.5)
Therapist	63 (24.3)
Professional	32 (12.4)
Adjunct	24 (9.3)
Admin	14 (5.4)
Lay person	3 (1.2)
Number of function (N=259), mean (SD)	8.4 (4.5)
Function type (N=259), n (%)	
Intervention	246 (95.0)
Execution	214 (82.6)
Social	175 (67.6)
Monitoring	103 (39.8)
Planning	22 (8.5)
Sophistication (N=259), mean (SD)	1.6 (0.6)
Intervention	1.5 (0.8)
Execution	1.7 (0.9)
Social	1.5 (0.9)
Monitoring	2.1 (1.1)
Planning	1.8 (1.0)

[a]Conducted on systems instead of versions.

Information and Communication Technology Platforms

The World Wide Web was the most frequently employed platform, with 75.7% (196/259) of the systems providing functionality on the Web and 6.2% (16/259) of the systems providing responsive website content that could also be displayed appropriately on mobile phones. Emails were sent or received in 43.2% (112/259) of systems. Following email,

telephone (20.5%, 53/259) and text messages (6.6%, 17/259) were frequently used to reach out to users. Only 1.9% (5/259) of the systems made use of storage media, such as CD and DVD and just as few exhibited technologies such as virtual agents (1.9%, 5/259), virtual reality (0.8%, 2/259), or connected to social media services (2.3%, 6/259).

Guidance

E-health software can include various types of human guidance or be entirely autonomous. Approximately half of all systems classified as the latter (47.5%, 123/259). In the remaining systems, guidance was mostly provided by the health care professionals. These were therapists in 24.3% (63/259) of cases and practitioners of related professions, such as coaches, nurses, social workers, or clinical psychology students in 12.4% (32/259) of cases. Less than 10% (24/259) of guided systems were offered as adjunct systems, that is, systems that support face-to-face therapy. A total of 5.4% (14/259) of systems were supported by technicians and other administrators, and only 1.2% (3/259) of systems asked for support by a layperson, typically a peer, friend, or family member of the user.

Size and Functionality

In terms of size, the average system offered 8 functions (Mdn=8), with a range from 1 to 21. Furthermore, systems had, on average, 6 modules (Mdn=6) and an intended usage duration of slightly less than 9 weeks (Mdn=8). Although nearly all software contained some intervention functions (95.0%, 246/259) and some support functions (91.5%, 237/259), the four support function types were not equally represented. A total of 82.6% (214/259) of systems included execution support, such as reminders via text message. Social support functionality was provided by 67.6% (175/259) of systems. This was either *direct*, whereby the user communicated with a human, or *indirect*, whereby the user could, for example, see that other people had performed the program before them. The least represented support function type (8.5%, 22/259) was planning support. A typical example of a planning support function was setting up a treatment schedule at the outset of the intervention. Within systems, intervention functions were dominant: systems contained, on average, twice as many intervention functions as support functions (V=15,079, P<.001, r=0.09). In addition, unguided and guided systems differed in their composition, with the former only having 63% of the number of functions of the latter ($F_{1,233}$=51.34, P<.001, R^2=0.18). This effect showed for both intervention (U=3467, P<.001, r=0.41) and support (U=3839.5, P<.001, r=0.28) functions.

Therapeutic Aspects

Although the literature search and filtering focused on systems aiming to reduce depressive symptoms, only 69.9% (181/259) of the identified software targeted depression exclusively. A total of 9.3% (24/259) of these specifically targeted users with a comorbid physical illness (eg, cancer, multiple sclerosis, and diabetes). A few systems supported comorbidities in general (nonspecific, 2.7%, 7/259). Of all systems, 16.6% (43/259) also considered anxiety. However, other mental comorbidities were excluded from the reviewed literature, as they typically formed the primary treatment objective (eg, in systems targeting psychotic conditions and depression simultaneously).

The most prominently represented intervention functions, present in 78.9% (194/259) of systems, were unrelated to specific therapies, that is, they could be categorized with many or all different therapies, such as *learning to recognize one's own symptoms* or *preventing relapses*. A large percentage of software made use of behavioral (62.6%, 154/259), cognitive (58.9%, 145/259), and cognitive behavioral (50.4%, 124/259) functions. Taken together, techniques related to cognitive behavioral therapy (CBT) were present in 84.9% (209/259) systems. Techniques from psychodynamic approaches were rare (2.0%, 5/259), as were life reviewing or hypnosis techniques (together present in 2.8%, 7/259, denoted by *others* in Table 3). A total of 69.5% (180/259) of systems had the purpose of treating depression and 29.3% (76/259) of systems had the purpose of preventing it. Only 1.2% (3/259) of the systems aimed to support patients in maintaining a depression-free state. The system purpose was related to the therapeutic approach (χ^2_7=34.1, P<.001, Nagelkerke R^2=0.24). Systems with Positive Psychology techniques were more often intended for prevention than for treatment. This was not the case for systems with techniques from other therapies (Figure 3).

Similarly, guided systems (χ^2_1=10.0, P=.002, Nagelkerke R^2=0.05) were more often used in treatment systems (n_{guided}=105, $n_{unguided}$=75), whereas unguided ones were used more in prevention (n_{guided}=28, $n_{unguided}$=48).

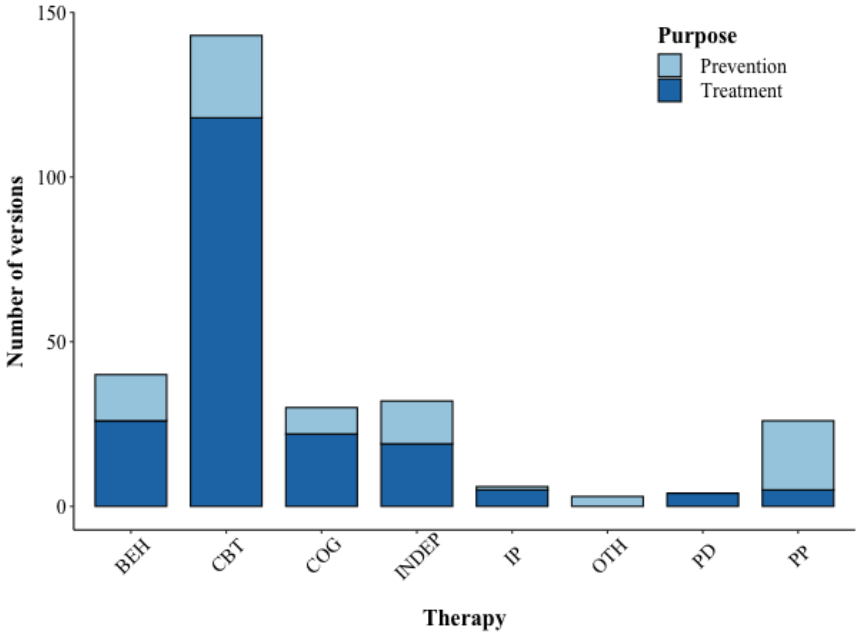
Table 3. The distributions over therapy-related key attributes of depression support system versions.

Therapy	Value
Comorbidity (N=259), n (%)	
None	181 (69.9)
Anxiety	43 (16.6)
Physical	24 (9.3)
Nonspecific	7 (2.7)
Addiction[a]	5 (1.9)
Insomia[b]	2 (0.8)
Purpose (N=259), n (%)	
Treat	180 (69.5)
Prevent	76 (29.3)
After-care	3 (1.2)
Duration (weeks; N=210), mean (SD)	8.7 (9.1)
Number of modules (N=218), mean (SD)	5.9 (3.5)
Therapy class (N=259), n (%)	
Independent	194 (78.9)
Behavioral	154 (62.6)
Cognitive	145 (58.9)
Cognitive behavioral therapy	124 (50.4)
Interpersonal	43 (17.5)
Positive psychology	43 (17.5)
Psychodynamic	7 (2.0)
Other	5 (2.8)

[a]Addiction is separated from physical illness, as it can be regarded as both a physical and a mental illness.

[b]Insomnia is separated from physical illness, as insomnia is also a symptom of depression.

Figure 3. The number of versions with the purpose of preventing or treating depression per therapy. A detailed list of the therapy subtypes for each of the therapy categories listed here can be found in Multimedia Appendix 5. BEH: behavioral therapy, CBT: cognitive behavioral therapy, COG: cognitive therapy, INDEP: independent of specific therapeutic theory, IP: interpersonal therapy, OTH: other, PD: psychodynamic therapy, PP: positive psychology.



Evaluation

Systems were often evaluated only once with end users (86.9%, 225/259) and, for the largest part, in comparative trials (77.2%, 200/259). In controlled trials, attention control (41.7%, 73/175) and waitlist (39.4%, 69/175) were similarly common, whereas treatment as usual (28.6%, 50/175) was less frequent (Table 4). In total, 72.2% (187/259) of systems were evaluated in controlled trials. Multimedia Appendix 7 comprises two tables ranking systems according to the number of evaluations and the total number of participants who participated in these studies.

Although 21 different measures assessed depressive symptomatology across studies, the most frequent by far were the Patient Health Questionnaire [26], Beck's Depression Inventory [27], and the Center for Epidemiological Studies Depression Scale [28]. An additional 11 measures were depression related, determining such things as fatigue, rumination, stress, or quality of life. Finally, 12.0% (31/259) of systems were evaluated in studies having primary outcomes other than depression, such as usability.

Table 4. The distributions over evaluation-related key attributes of depression support system versions.

Evaluation	Value
Number of studies, mean (SD)	1.2 (0.9)
Quality (N=259), n (%)	
Comparative	200 (77.2)
Noncomparative	74 (28.6)
Control group types (N=259), n (%)	
Attention controlled	73 (41.7)
Waitlist	69 (39.4)
TAU[a]	50 (28.6)
Measures (N=259), n (%)	
PHQ[b]	90 (34.7)
BDI[c]	74 (28.6)
CES-D[d]	65 (22.0)
Other depression measure	57 (25.1)
Nondepression measure	31 (12.0)

[a]TAU: treatment as usual.

[b]PHQ: Patient Health Questionnaire.

[c]BDI: Beck Depression Inventory.

[d]CES-D: Center for Epidemiological Studies Depression.

Description of a Fictional, Prototypical System

For illustration purposes, we outline here a fictional, prototypical depression treatment system by combining insights from the qualitative reading of the articles and the quantitative analyses. This is intended to serve as a narrative description of the taxonomy provided in Multimedia Appendix 5. However, it must be noted that this is a simplification and much variation exists among the systems. A prototypical system takes a CBT approach and might comprise 6 modules, one of which is released every week. The modules can be accessed on a website. The participant is made aware of the presence of a new module via email; thus, the participant is reminded to adhere to the treatment. Modules might cover topics such as activity scheduling, learning to detect automatic thoughts, cognitive restructuring, problem solving, psychoeducation concerning depression and the therapeutic approach, and relapse prevention. Each module comes with exercises that are submitted to be checked by a therapist or similar, who again provides feedback via email. The website might include a small calendar application for the purposes of activity scheduling and a diary application for the purposes of thought recording. In these applications, the user can enter and save information. Once a week, the participant is asked to complete a depression scale, and the therapist is notified if suicidal ideation is detected. The remaining questions are averaged and presented to the user as a mood graph on the landing page. This sketched system would have an average eHDTS score of around 2. For each of the eHDTS levels, a similar, fictional description of possible functions scoring at this level can be found in Multimedia Appendix 8. This is intended to provide a more concise and tangible description than Multimedia Appendix 6 can and to further concretize the taxonomy presented in Multimedia Appendix 3.

Technological Sophistication

Systems

The average system comprised, to a large extent, functions providing information to the user without collecting and interpreting information from the user. This is further detailed in Figure 4. Almost all interventions had the majority of their

functions delivered through technology, that is, hardly any system scored below 1 on technological sophistication. However, only 21.1% (28/133) of systems had a sophistication level above 2, indicating that they were responsive to activities and information coming from the user. These systems comprised, for the most part, interventions inspired by CBT or closely related therapies. In fact, CBT systems lead the list of the most technologically advanced systems, even when adjusting for the number of functions (Table 5). The top two systems in both rankings are Help4Mood [29] and Deprexis [30]. The latter is a commercial system aiming to mimic the structure of face-to-face CBT therapy, whereas the former is a self-monitoring system that includes a virtual conversational agent. Both presented high-tech solutions according to the eHDTS scale, as they adapted the intervention to the users' indicated interests and needs (Deprexis) or to the self-monitoring data from users (Help4Mood). To allow researchers to compare their own system, Multimedia Appendix 9 provides the eHDTS score per cumulative percentage decile of systems for both the weighted and unweighted system means. That is, when knowing the average weighted or unweighted eHDTS score of their system, researchers can use the table to determine which decile of systems their system scores at, below, or above.

Technological sophistication was not linked to the number of functions (r_{257}=0.01, P=.83), the system purpose (χ^2_1=0.2, P=.69), or guidance (χ^2_1=3.0, P=.08). However, it did relate to the evaluation quality (χ^2_1=6.1, P=.01, Nagelkerke R^2=0.03). More technologically sophisticated systems were less likely (OR 0.59) to have been evaluated in comparative trials than less technologically sophisticated systems. Furthermore, when regarding specifically randomized controlled trials (RCTs), we found that 80.8% (139/172) of RCTs evaluate systems that score below data entry level on average, with the respective percentage of RCTs per eHDTS interval being the following: [0,1)—4%; [1,2)—77%; [2,3)—16%; and [3,4)—3%.

Figure 4. Cumulative density plot of all systems over the e-mental Health Degree of Technological Sophistication (eHDTS) scale. This analysis was conducted on the unweighted average of technological sophistication of the systems. Labeled dots show the highest scoring system within a specific therapy, as indicated by the label. BEH: behavioral therapy, CBT: cognitive behavioral therapy; COG; cognitive therapy; eHDTS: e-mental Health Degree of Technological Sophistication; INDEP: independent of specific therapeutic theory; IP: interpersonal therapy; Oth: other; PD: psychodynamic therapy; PP; positive psychology.

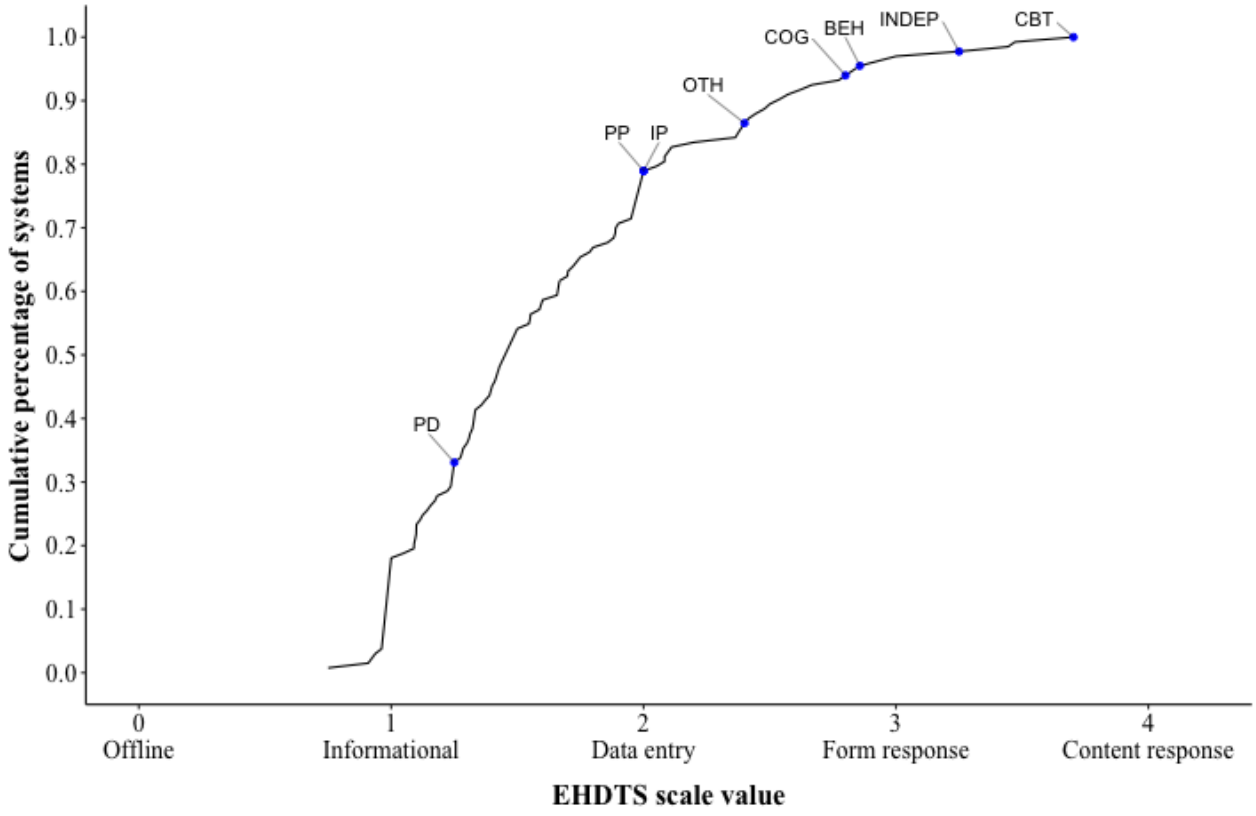


Table 5. Ranking of the 10 systems with the highest degree of technological sophistication in the database, first based on average e-mental Health Degree of Technological Sophistication (eHDTS) score (M) and then based on a weighted eHDTS score (M_w), trading off eHDTS against the number of functions in a system. The analyses were conducted on the basis of systems rather than versions. We advise some caution in taking this table at face value, as it is based on the aggregated eHDTS scores with some of the scales only having moderate interrater agreement.

Rank	Unweighted				Weighted				
	System	Therapy	n_f [a]	M [b]	System	Therapy	n_f	M	M_w [c,d]
1	Help4Mood [29]	CBT[e]	13.5	3.70	Help4Mood [29]	CBT	13.5	3.70	2.31
2	Deprexis [30]	CBT	14.3	3.47	Deprexis [30]	CBT	14.3	3.47	2.31
3	MOSS App [31]	CBT	9	3.44	Buhrman [32]	CBT	20	1.95	1.85
4	Ahmedani [33]	MI[f], CBT	4	3.25	Building a Meaningful Life through BA[g] [34]	BA	20	1.90	1.81
5	DCAT ATA [35]	SM[h]	5	3.00	Shamekhi [36]	MFN[i]	13	3.00	1.80
6	Shamekhi [36]	MFN	13	3.00	Living to the full [37]	ACT[j]	14.5	2.62	1.77
7	Panoply [38]	CBT	7	2.86	MindBalance [39]	CBT	14	2.57	1.67
8	MyPAA [40]	PhA[k]	7	2.86	Space from Depression [41]	CBT	15	2.20	1.54
9	EVO [42]	CCT[l]	5	2.80	Mobilyze! [43]	BA	13	2.54	1.52
10	Daybuilder [44]	SM	6.5	2.77	MOSS App [31]	CBT	9	3.44	1.38

[a]n_f: number of functions.

[b]M: unweighted average.

[c]M_w: weighted average.

[d]To obtain the weighted average (M_w), the unweighted average (M) is weighted with the feature scaled number of functions (nf): Mw=M(nf – min(nf))/(max(nf)-min(nf)), with min(nf)=1 and max(nf)=21.

[e]CBT: cognitive behavioral therapy.

[f]MI: motivational interviewing.

[g]BA: behavioral activation.

[h]SM: symptom monitoring.

[i]MFN: mindfulness.

[j]ACT: acceptance and commitment therapy

[k]PhA: physical activity.

[l]CCT: cognitive control training.

Functions

Support functions (mean 1.73, SD 1.06) scored higher in technological sophistication than intervention functions (mean 1.43, SD 0.88), although this effect was small ($F_{1,1903}$=38.11, P<.001, Level1 R^2=0.03). An equally small effect was observed while comparing the 4 types of support functions on their technological sophistication ($F_{3,619}$=8.46, P<.001, Level1 R^2=0.04). Monitoring support functions had the highest average degree of technological sophistication (Table 2). This indicates that monitoring functions were mostly technologically sophisticated to the extent that they reported data back to the user, but they neither interpreted data nor used the data to adapt the intervention. Social support and intervention functions ranked the lowest in terms of technological sophistication (Table 2). In social support, the score translates to technology being typically either used to simply provide contact information to the user or to serve as a communication medium between human support and user. Intervention functions often took an informational form, possibly with a limited amount of interactivity, for example, clicking through pages or filling in a Web-based diary.

The most frequently implemented support functions were execution support pertaining to the management of user progress and risk, triggers, indirect social support, professional direct social support, and symptom monitoring (Figure 5). However, only management execution support and indirect social support were present at least once in systems of all different therapies. A barely implemented function type was planning support. Shifting the focus to intervention functions, most stem from CBT or related therapies or are independent of a specific therapeutic framework. CBT systems clearly dominate the field, with most of the different function types being implemented in numerous such systems (Figure 5). Yet, the average technological sophistication of functions (Figure 6) was not related to how frequently they were implemented (r_{154}=0.12, P=.12). Thus, functions that are often implemented are neither more nor less technologically sophisticated, on average, than functions that are rarely implemented. However, the more often a function was implemented, the more often at least 1 of these implementations was responsive to interaction activity of the

user, for example, time spent on platform, or even to the content of information provided by the user (r_{154}=0.43, P<.001). For interested readers, Multimedia Appendix 10 finally also demonstrates that nearly all of the different functions were implemented in a highly sophisticated manner in at least one system.

Figure 5. Heatmap of the frequency with which a specific type of function was implemented in a therapy across all systems of that therapy. BEH: behavioral therapy, CBT: cognitive behavioral therapy, COG: cognitive therapy, INDEP: independent of specific therapeutic theory, IP: interpersonal therapy, OTH: other, PD: psychodynamic therapy, PP: positive psychology.

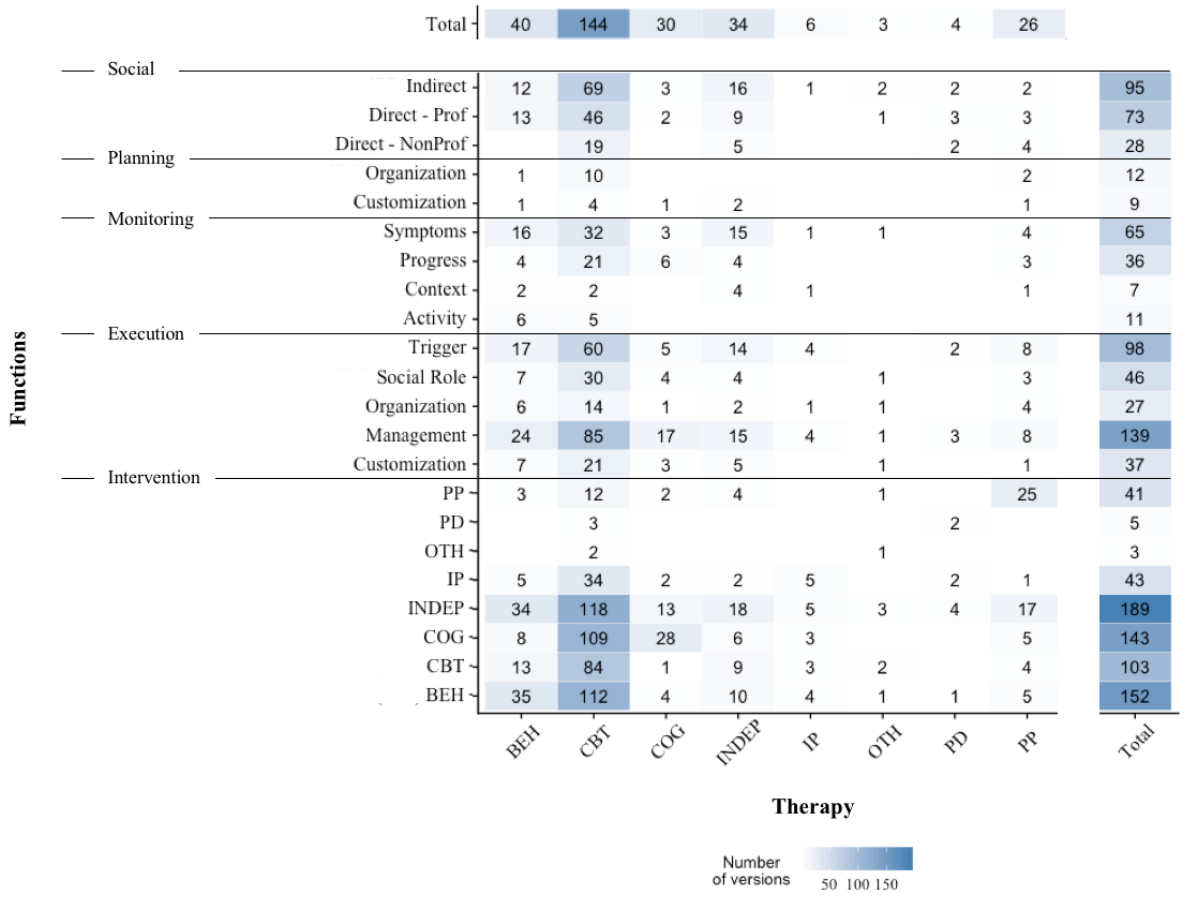
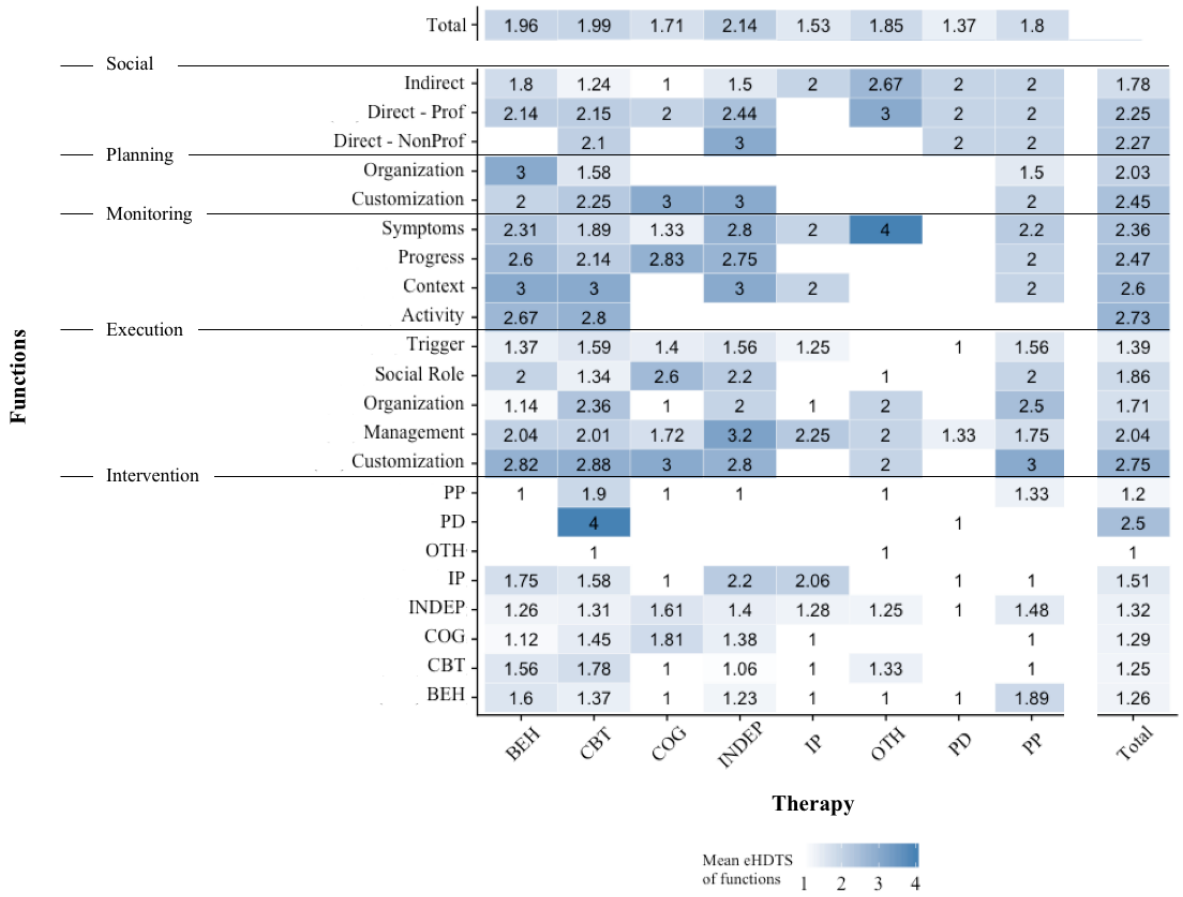


Figure 6. Heatmap of the average degree of technological sophistication per function type and therapy. BEH: behavioral therapy, CBT: cognitive behavioral therapy, COG: cognitive therapy, INDEP: independent of specific therapeutic theory, IP: interpersonal therapy, OTH: other, PD: psychodynamic therapy, PP: positive psychology.



Developments Over Time

In the past 2 decades, the field of e-mental health for depression has seen marked growth, with 5 times as many systems developed in 2014 as in 2000 (Figure 7). As several years typically lie between development and the publication of study outcomes, less emphasis may be given to numbers after 2014. The figure also demonstrates that systems were being reused and extended to a substantial degree only from approximately 2009 onward. This is further supported, when examining systems with at least five versions more closely (Figure 8). Only MoodGYM had evolved multiple versions before 2009. Different versions developed within the same year are an indication that they were created for the same study, often differing in only 1 function as an experimental manipulation.

Despite growth in the field in general, systems seemed to neither get larger ($F_{1,257}$=0.25, P=.62) nor more sophisticated ($F_{1,257}$=1.88, P=.17) with time. Within systems, growth was observed across versions, with each new version of a system having half of a function more than the previous one (b=0.50, $F_{1,125}$=11.60, P<.001, Level1 R^2=0.06). However, technological sophistication seemed to remain the same ($F_{1,125}$=1.96, P=.16). Finally, the evaluation quality showed no relationship with the version number ($F_{1,136}$=0.07, P=.79). Later versions therefore appeared to be no more or less frequently evaluated in comparative trials than earlier ones.

Figure 7. The number of systems and versions developed per year between 2000 and 2016.

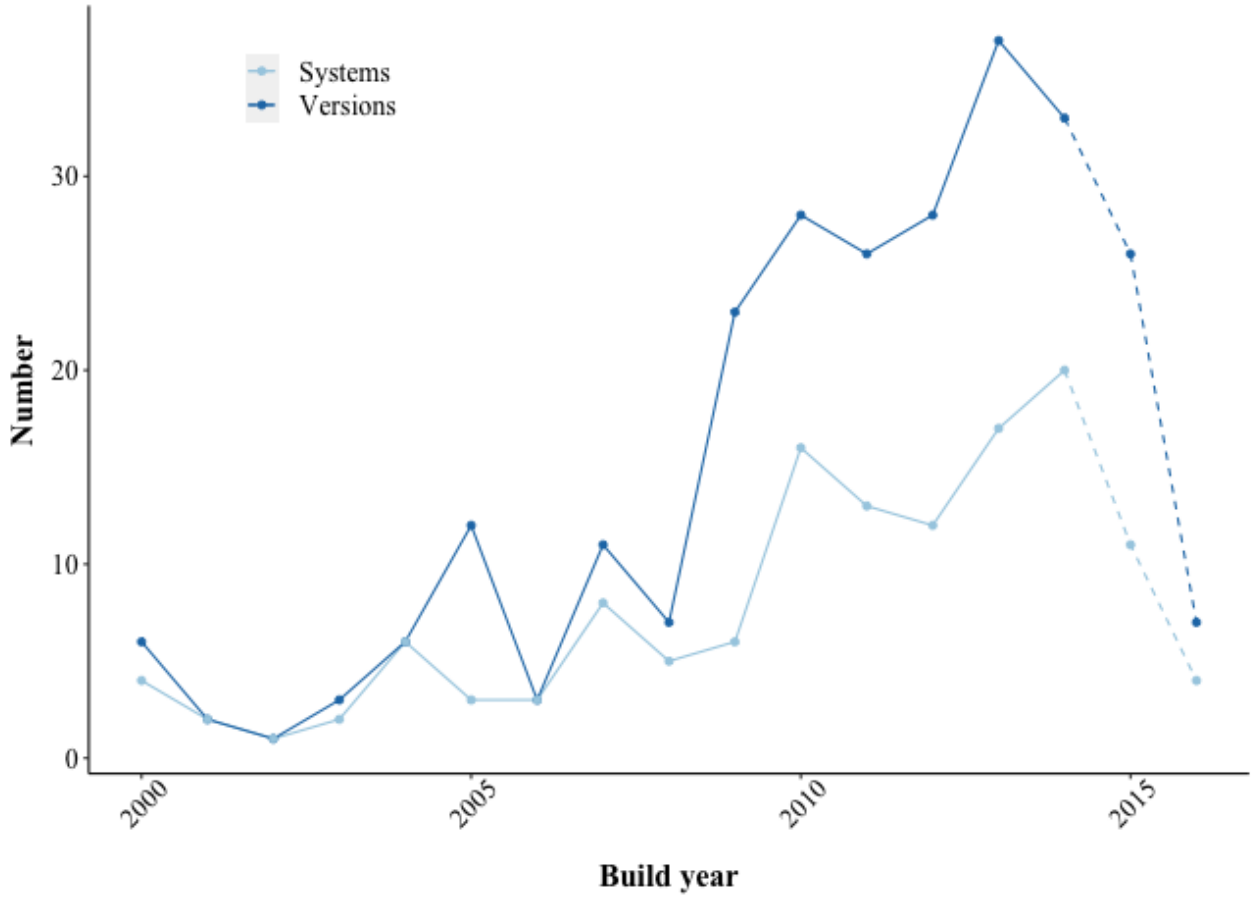
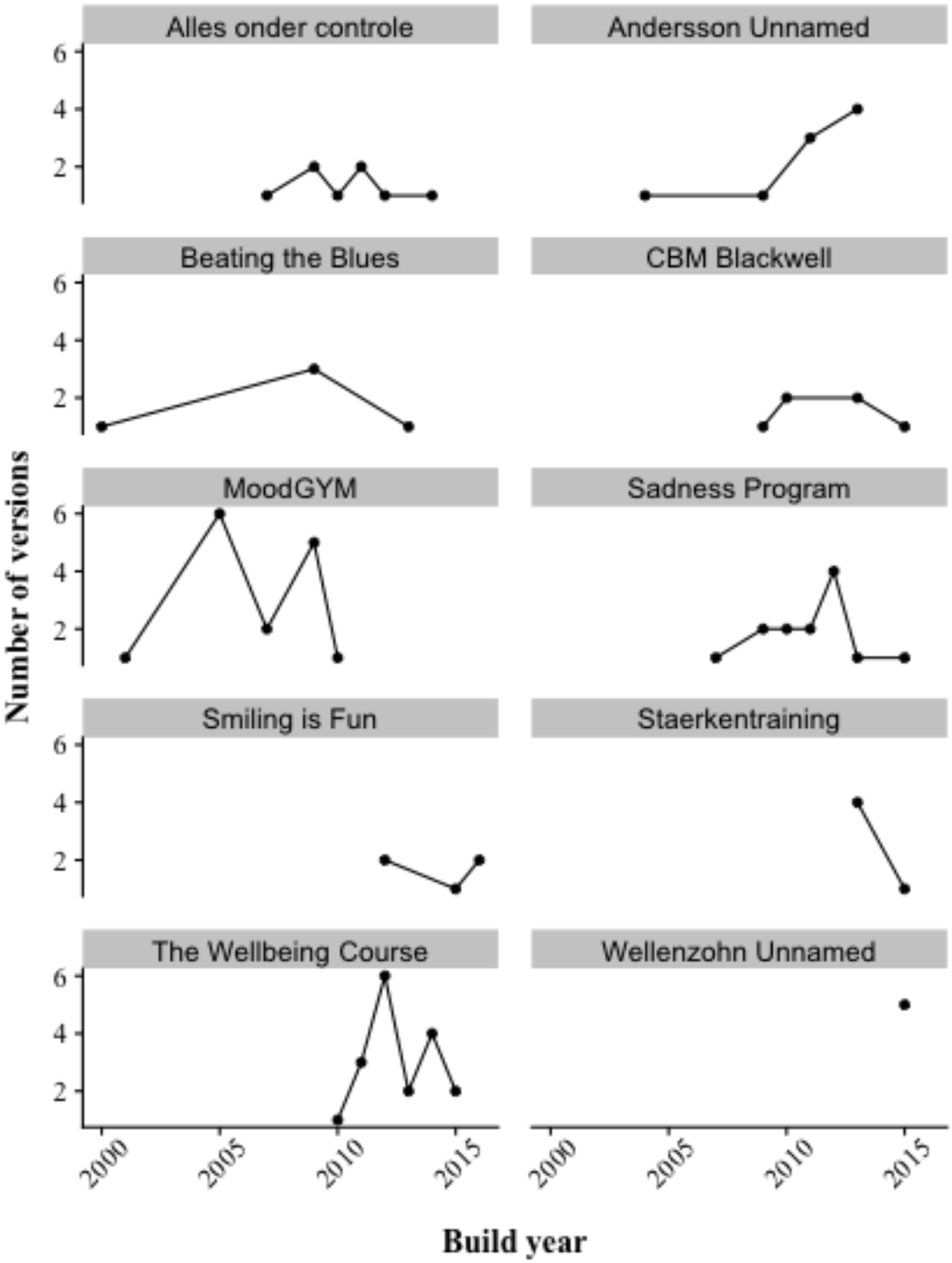


Figure 8. The number of versions developed per year between 2000 and 2016 for the ten systems having five or more versions.



Discussion

Principal Findings

Some limitations should be kept in mind when considering implications of the results. The first pertains to the coding of technological sophistication. Namely, the one-dimensional nature of the eHDTS scales can limit them in covering the full extent of the degree to which they reflect how technologized a function is. In monitoring support functions, for example, the scale captures how the system deals with the collected information but not how the data are obtained in the first place. Thus, whether monitoring data are collected via self-report or sensing does not influence the level of technological sophistication. However, as sensors and data analysis methods are becoming increasingly reliable, sensing will likely begin to play a crucial role in more automated, that is, more technologically advanced, systems [45]. In addition to

monitoring, this is particularly to be expected in diagnosis and assessment systems [46,47], which we have excluded in this review. Thus, in the future and especially when wishing to also study such systems, the manner in which data are collected should receive more attention in the monitoring scale. An additional point to consider when interpreting the results is the moderate reliability of some high-inference attributes. Although this is a limitation that might influence more detailed findings, such as the exact ranking of the systems according to their eHDTS score, we do not expect it to substantially affect the larger patterns found. However, by double coding samples, we have insight into the reliability of the estimates. In selecting a sample size for double coding, we have aimed for a 10% margin of error for the reliability estimates, as suggested in [48] (see [11] for 95% CI information for each estimate). Two final limitations concern the scope of the reviewed systems, as well as the scope of the reviewed functions. Systems developed for children and adolescents, women with depression during or following pregnancy, and those with comorbid psychiatric conditions, as well as systems developed before 2000, were excluded. In addition, we did not consider commercial systems that are not reported in the scientific literature. How well our findings generalize to these types of systems is therefore open to further investigation. As far as the scope of the reviewed functions is concerned, functions pertaining to data security or the integration of the system with existing health software were not covered. Such aspects of the interventions were typically not found to be reported in the publications. As data security is becoming an important concern of software development and usage, we see a need for more consistent system reporting guidelines and an opportunity for reviews of future systems to subsequently investigate such functionality. In spite of these limitations, the outcomes of the analyses highlight some of the challenges and opportunities for the field of e-mental health for depression.

First, no clear progress in terms of system sophistication was observed between 2000 and 2017, within or across systems. A possible challenge for progress might lie in the short-term approach to system development in the field. In a long-term approach, multiple versions with substantial changes in functionality could be expected. Early versions would be tried in pilot studies, improved, and only eventually tested in an RCT. However, this is not what we found. Despite often proving effective in RCTs, two-thirds of the systems are not evolved and retested (eg, [49-51]). In addition, in systems that do have multiple versions, systems are often extended only by a function for hypothesis testing among versions, and versions do not differ in technological sophistication. Finally, there was also no association between the evaluation level and the version number for systems that had more than 1 version.

Another challenge for the field is posed by the spread in technological sophistication. Our analyses confirmed what has been hinted at in previous reviews and meta-analyses [2,6,7]: systems developed within a research context vary in their implementation and in their technological sophistication. By and large, they are not very technologically advanced, and those systems that are mostly informational in nature account for 81% of what is evaluated in RCTs. Only approximately one-fifth of the systems have a substantial amount of functions that are responsive to input from the user. These differences in technological realization have, thus far, been neglected in literature syntheses taking a clinical psychology perspective. For example, two effects identified in such syntheses are that both adherence and effect size appear to increase with higher levels of human guidance (no guidance vs administrative guidance vs therapist guidance) [3]. Although this has been hypothesized to be linked to missing therapeutic alliance or accountability, our results indicate another possibility. We found the lack of guidance to be neither compensated with more content or technological support nor with a more responsive and, thus potentially more engaging, system. It is therefore possible that guidance plays a role, especially when systems are not very responsive. As, according to our analyses, this applies to approximately 80% of the systems, the results of meta-analyses over all systems may not generalize to more technologically advanced solutions. This notion finds some support in a system-specific meta-analysis of the Deprexis system [52], which ranked second in our ranking of the most technologically advanced systems. Across different studies with Deprexis, dropout ranged from 6% to 50%, contrasting with the average dropout rate of 74% found for other unguided systems in general [3]. Furthermore, it was not only observed that unguided Deprexis had an average effect size across trials comparable with that of other systems, including administrative guidance [3], but also that adding guidance did not influence the magnitude of the effect. However, it must also be emphasized at this point that the potential of more technologically advanced systems leading to higher adherence is merely a hypothesis that is in need of further investigation.

Aside from these challenges, we also see opportunities. Systems developed for depression, to date, are hardly making use of the full bandwidth of available technology. In fact, empirically evaluated systems are mostly delivered on the World Wide Web. Only a very few take a mobile form as either native apps or cross-platform Web applications. This is surprising considering that smartphones became a ubiquitous and highly used technology approximately mid-way of the examined time period. In a review from 2015 on the state of the app marketplace for depression apps, 82 apps had been identified for the treatment of depression [53]. A later review (2017) found that only 5 apps for depression treatment had been empirically evaluated in effectiveness trials [54]. Therefore, an abundance of apps exists, but most apps are commercial, and few have been scientifically studied. However, the empirically evaluated apps included in this review fared well in technological sophistication, such as Mobilyze! [43] and Mobile Sensing and Support (MOSS) [31]. Both apps attempt to learn how to provide context-sensitive interventions on the basis of phone sensor readings. The former uses models trained before delivering the interventions, whereas the latter continuously learns user preferences as it intervenes. In addition to mobile apps, there are several other underexplored innovative technologies, such as social media, conversational agents, and virtual reality. Yet, where these were implemented, some technologically interesting solutions emerged. In social media systems, Panoply [38] can be considered a technological forerunner. It integrates social networking between Panoply users and crowdsourcing from

Amazon Mechanical Turk to ensure high-quality content, both in terms of users' thought-recording posts and in terms of responses to these posts. Woebot [55], a fully autonomous chatbot provided on social media, was developed after our search; therefore, it was not included in the analyses. Through short, daily conversations using Facebook instant messenger, Woebot continuously checks in with users and tailors short intervening information and empathic replies to their reported mood. Finally, a creative attempt to alleviate depression is presented by the only virtual reality system that we found [56]. Users are first asked to comfort a virtual avatar with the embodiment of a child. They then take on the perspective of this child in virtual reality to hear their own comforting words said back to them, with the effect of increasing their self-compassion. However, innovative technology solutions, such as the ones mentioned, are scarce. Thus, there still are many opportunities for the field to explore such directions.

Conclusions

The e-mental health field, focusing specifically on the treatment and prevention of depressive disorders, is large and consist of a very active research community, as evidenced by the vast body of literature that could be identified for this study. In line with our research questions, three main conclusions can be drawn. First, although the system landscape is overall varied, there are clear trends: three quarters of the systems implement therapeutic techniques related to CBT, three quarters are delivered on the World Wide Web, and three quarters have been evaluated in comparative trials. Second, most systems do not get close to the full technological potential of e-mental health. However, some do get close. On the level of functions, we have further found that nearly all functions have been implemented in a responsive manner in at least one system, showing that the high end of the scale is obtainable across the board. Third, there appears to be no clear technological development across systems between 2000 and 2017. Furthermore, within systems that have multiple versions, a small increase in size with each new version showed, but it was not the case in technological sophistication. Consequently, it can be argued that, from a technological perspective, there is still room for improvement. Future research investigating the relationship between software implementation and clinical outcomes will need to show whether such improvement is beneficial and cost-efficient with regard to development and maintenance.

To conclude, the scientific contribution of this research is its provision of a comprehensive overview of the technological state of the art of e-mental health systems for the prevention and treatment of adult major depressive disorder, developed and studied since the year 2000. This is further accompanied by EHealth4MDD, an open-access database containing all extracted and coded information from the literature used in this writing. Together, the review and database are intended to serve as inspiration for the development of new systems on the one hand and as facilitators for the study of hypotheses related to system composition, on the other hand.

Acknowledgments

This work has been funded by the 4TU research center Humans & Technology project (Systems for Smart Social Spaces for Living Well: S4).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search terms that were used in retrieving primary articles from the databases Scopus, PubMed, and Web of Science are shown. Columns are combined with logical ANDs, whereas cells within columns are combined with logical ORs. The first 2 columns were searched for within titles and keywords; the third and fourth column were searched for within abstracts. Terms from the Exclude column were not allowed to appear within titles only. The first and second column include in italics the Medical Subject Heading terms that were used in addition to the regular search terms in PubMed. Finally, where possible, wildcards were used to expand the search terms with * denoting any string, including the empty one, and $ denoting any string of length 1 or the empty string.
[PDF File (Adobe PDF File), 265 KB - jmir_v22i1e12599_app1.pdf]

Multimedia Appendix 2
List of all publications that were included in this review.
[PDF File (Adobe PDF File), 324 KB - jmir_v22i1e12599_app2.pdf]

Multimedia Appendix 3
Taxonomy of software systems for depression. This was inspired by the one for conversational agents, presented in Montenegro JL, da Costa CA, and da Rosa Righi R. "Survey of conversational agents in health." Expert Systems with Applications (2019). It is important to note that this is not an exact graphical representation of all concepts in the database but is intended as an illustration of the most descriptive attributes of software systems for depression. For readers interested in an exact graphical representation of the database, we refer to the SQL schema diagram on the EHealth4MDD (database) website.

[PDF File (Adobe PDF File), 32 KB - jmir_v22i1e12599_app3.pdf]

Multimedia Appendix 4
Categorization of support functions at level 0 (L0), level 1 (L1), and level 2 (L2). Customization as an Execution function type means that the system or intervention could be altered throughout the usage period according to the user's preferences, while as a Planning function type, customization was only possible at the start of the intervention. The difference between Management and Organization within the Execution type is that Organization pertains to management of aspects of the specific system and intervention, whereas Management pertains to dealing with higher-level problems or aspects.
[PDF File (Adobe PDF File), 39 KB - jmir_v22i1e12599_app4.pdf]

Multimedia Appendix 5
The categorization of therapeutic frameworks into therapies (L1). An example function at level 0 (L0) is provided for each therapeutic framework. Some of the therapies were mentioned by authors as having influenced the design of the system, but using our classification, no intervention functions pertaining to the therapy could be found. Therefore, no example can be given. This does not mean that no functionality reflecting the therapy was implemented, for example, a symptom monitoring approach might well result in functionality to aid in the monitoring of symptoms. However, with our classification, this would be classified as a monitoring support function rather than as an intervention function. Similarly, influences from Social Cognitive Theory may have found their way into the system in the form of vignettes, which we classify as indirect social support functions rather than intervention functions.
[PDF File (Adobe PDF File), 39 KB - jmir_v22i1e12599_app5.pdf]

Multimedia Appendix 6
Part of the EHealth4MDD (database) website, detailing the 5 different subscales of e-mental Health Degree of Technological Sophistication. Exact level definitions for each of the scales and an example function are provided.
[PDF File (Adobe PDF File), 138 KB - jmir_v22i1e12599_app6.pdf]

Multimedia Appendix 7
A total of 2 tables showing the ranking of the 133 systems contained in the database by evidence base, as far as this has been recorded in the database. The first table shows the evidence base of systems evaluated in comparative trials (such that are randomized, controlled, or both), whereas the second table shows the evidence base of systems evaluated in noncomparative trials (single-group trials). Both tables are sorted first on the number of evaluations, on the number of participants recruited to take part in the study (sum over all study arms), and the number of participants who completed the study (not including follow-up). For readers interested in more information, the system key, as denoting systems in the database, is provided. This should allow for easy querying of associated versions, authors, and articles, to name only a few things.
[PDF File (Adobe PDF File), 337 KB - jmir_v22i1e12599_app7.pdf]

Multimedia Appendix 8
Table providing a description of some possible functions that a system at each of the e-mental Health Degree of Technological Sophistication (eHDTS) score levels might have. These descriptions are fictional, and eHDTS scores at the system level are averages.
[PDF File (Adobe PDF File), 195 KB - jmir_v22i1e12599_app8.pdf]

Multimedia Appendix 9
Deciles and their corresponding scale values for the weighted and unweighted scale, for example, 50% of systems have an average technological sophistication of 1.5 or lower. The weighted column takes into account the number of functions that a system implements.
[PDF File (Adobe PDF File), 46 KB - jmir_v22i1e12599_app9.pdf]

Multimedia Appendix 10
Heatmap of the maximum degree of technological sophistication per function type and therapy. This gives insight into the technological state of the art of each function type and therapy.
[PNG File , 139 KB - jmir_v22i1e12599_app10.png]

References

1. World Federation for Mental Health. World Health Organization. 2012. Depression: A Global Crisis URL: https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf [accessed 2019-11-22]

2. Schröder J, Berger T, Westermann S, Klein JP, Moritz S. Internet interventions for depression: new developments. Dialogues Clin Neurosci 2016 Jun;18(2):203-212 [FREE Full text] [Medline: 27489460]
3. 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]
4. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. BMC Psychiatry 2011 Aug 12;11:131 [FREE Full text] [doi: 10.1186/1471-244X-11-131] [Medline: 21838902]
5. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
6. Wildeboer G, Kelders SM, van Gemert-Pijnen JE. The relationship between persuasive technology principles, adherence and effect of web-based interventions for mental health: a meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
7. 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]
8. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
9. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
10. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26(2-4):109-160. [doi: 10.1080/15228830802094429]
11. Burger F, Neerincx MA, Brinkman WP. EHealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. Cyberpsychol Behav Soc Netw 2018;21(9):18-24 [FREE Full text]
12. EHealth4MDD database. 2018. URL: http://insyprojects.ewi.tudelft.nl:8888/ [accessed 2019-11-26]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Microsoft Patterns & Practices Team. Microsoft Application Architecture Guide. USA: Microsoft Press; 2009.
15. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):96. [doi: 10.17705/1cais.02428]
16. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
17. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Burlington, MA, USA: Morgan Kaufmann; 2002.
18. Wendel S. Designing For Behavior Change: Applying Psychology And Behavioral Economics. Boston, MA, USA: O'Reilly Media; 2013.
19. Fassbinder E, Klein JP, Sipos V, Schweiger U. Therapie-Tools Depression. Weinheim, Germany: Beltz Verlag; 2015.
20. Grant MJ, Button CM, Snook B. An evaluation of interrater reliability measures on binary tasks using d-Prime. Appl Psychol Meas 2017 Jun;41(4):264-276 [FREE Full text] [doi: 10.1177/0146621616684584] [Medline: 29881092]
21. Lemma A, Fonagy P. Feasibility study of a psychodynamic online group intervention for depression. Psychoanal Psychol 2013;30(3):367-380. [doi: 10.1037/a0033239]
22. Currie SL, McGrath PJ, Day V. Development and usability of an online CBT program for symptoms of moderate depression, anxiety, and stress in post-secondary students. Comput Human Behav 2010;26(6):1419-1426. [doi: 10.1016/j.chb.2010.04.020]
23. Burger F, Brinkman WP, Neerincx MA. 4TU - DATASETS. 2019. Analysis of technological state of the art in e-mental health for the prevention and treatment of depression URL: https://data.4tu.nl/repository/uuid:623b37bc-3ea7-48e3-ab09-575aa3b1b9ac [accessed 2019-11-26]
24. Nagelkerke NJ. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691-692. [doi: 10.1093/biomet/78.3.691]
25. Snijders TA, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London, UK: Sage Publications Ltd; 1999.
26. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. J Am Med Assoc 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
27. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4(6):561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]

28. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
29. Burton C, Tatar AS, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
30. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009 May 11;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
31. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
32. Buhrman M, Syk M, Burvall O, Hartig T, Gordh T, Andersson G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):504-516. [doi: 10.1097/AJP.0000000000000176] [Medline: 25380222]
33. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. Behav Med 2015;41(1):25-32. [doi: 10.1080/08964289.2013.867827] [Medline: 24313728]
34. Spates CR, Kalata AH, Ozeki S, Stanton CE, Peters S. Initial open trial of a computerized behavioral activation treatment for depression. Behav Modif 2013 May;37(3):259-297. [doi: 10.1177/0145445512455051] [Medline: 22987916]
35. Wu S, Vidyanti I, Liu P, Hawkins C, Ramirez M, Guterman J, et al. Patient-centered technological assessment and monitoring of depression for low-income patients. J Ambul Care Manage 2014;37(2):138-147 [FREE Full text] [doi: 10.1097/JAC.0000000000000027] [Medline: 24525531]
36. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting group medical visits with conversational agents for stress management behavior change. In: Proceedings of the International Conference on Persuasive Technology. 2017 Presented at: PERSUASIVE'17; April 4-6, 2017; Amsterdam, The Netherlands p. 55-67.
37. Kelders SM, Bohlmeijer ET, Pots WT, van Gemert-Pijnen JE. Comparing human and automated support for depression: fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]
38. Morris RR, Schueller SM, Picard RW. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J Med Internet Res 2015 Mar 30;17(3):e72 [FREE Full text] [doi: 10.2196/jmir.4167] [Medline: 25835472]
39. 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.: ACM; 2012 Presented at: CHI'12; May 5-10, 2012; Austin, TX, USA p. 1421-1430.
40. Rebar AL, Boles C, Burton NW, Duncan MJ, Short CE, Happell B, et al. Healthy mind, healthy body: a randomized trial testing the efficacy of a computer-tailored vs interactive web-based intervention for increasing physical activity and reducing depressive symptoms. Ment Health Phys Act 2016 Oct;11:29-37. [doi: 10.1016/j.mhpa.2016.08.001]
41. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
42. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. Depress Anxiety 2017 Jun;34(6):508-517 [FREE Full text] [doi: 10.1002/da.22588] [Medline: 28052513]
43. 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 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
44. Nørregaard LB, Løventoft PK, Frøkjær E, Lauritsen L, Olsson EC, Andersen L, et al. Patient Expectations and Experiences From a Clinical Study in Psychiatric Care Using a Self-monitoring System. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. New York, NY, USA: ACM; 2014 Presented at: NordiCHI'14; October 26 - 30, 2014; Helsinki, Finland p. 991-994. [doi: 10.1145/2639189.2670258]
45. Riva G, Banos R, Botella C, Gaggioli A, Wiederhold B. Personal health systems for mental health: the European projects. Stud Health Technol Inform 2011;163:496-502. [Medline: 21335846]
46. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
47. Liu Z, Hu B, Yan L, Wang T, Liu F, Li X. Detection of Depression in Speech. In: Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction. 2015 Presented at: ACII'15; September 21-24, 2015; Xi'an, China p. 743-747. [doi: 10.1109/acii.2015.7344652]
48. Gwet KL. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Gaithersburg, MD, USA: Advanced Analytics, LLC; 2010.

49. Hunkeler EM, Hargreaves WA, Fireman B, Terdiman J, Meresman JF, Porterfield Y, et al. A web-delivered care management and patient self-management program for recurrent depression: a randomized trial. Psychiatr Serv 2012 Nov;63(11):1063-1071. [doi: 10.1176/appi.ps.005332011] [Medline: 22983558]
50. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an Internet-based cognitive behavioral therapy (iCBT) program in Manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS One 2014;9(5):e97167 [FREE Full text] [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]
51. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
52. Twomey C, O'Reilly G, Meyer B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: a meta-analysis. Psychiatry Res 2017 Oct;256:371-377. [doi: 10.1016/j.psychres.2017.06.081] [Medline: 28686935]
53. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
54. van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
55. 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 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
56. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Jan;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]

Abbreviations

CBT: cognitive behavioral therapy
eHDTS: e-mental Health Degree of Technological Sophistication
ICT: information and communication technology
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 24.10.18; peer-reviewed by M Alcaniz Raya, S Ondersma, D MacIntyre; comments to author 21.03.19; revised version received 16.08.19; accepted 31.10.19; published 20.01.20.

Please cite as:
Burger F, Neerincx MA, Brinkman WP
Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review
J Med Internet Res 2020;22(1):e12599
URL: https://www.jmir.org/2020/1/e12599
doi:10.2196/12599
PMID:31958063

Original Paper

Associations Between Affective States and Sexual and Health Status Among Men Who Have Sex With Men in China: Exploratory Study Using Social Media Data

Zhi-Wei Zheng[1*], MS; Qing-Ling Yang[1*], MS; Zhong-Qi Liu[1], MS; Jia-Ling Qiu[1], MS; Jing Gu[1,2,3], PhD; Yuan-Tao Hao[1,2,3], PhD; Chao Song[4], PhD; Zhong-Wei Jia[5], PhD; Chun Hao[1,2,3], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Health Information Research Center, Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou, China

[3]Sun Yat-sen Global Health Institute, Institute of State Governance, Sun Yat-sen University, Guangzhou, China

[4]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

[5]National Institute on Drug Dependence, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Chun Hao, PhD
Department of Medical Statistics
School of Public Health
Sun Yat-sen University
No 74, Zhongshan 2nd Road
Guangzhou, 510080
China
Phone: 86 13922750299
Email: haochun@mail.sysu.edu.cn

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/3/e18135/

Abstract

Background: Affective states, including sentiment and emotion, are critical determinants of health. However, few studies among men who have sex with men (MSM) have examined sentiment and emotion specifically using real-time social media technologies. Moreover, the explorations on their associations with sexual and health status among MSM are limited.

Objective: This study aimed to understand and examine the associations of affective states with sexual behaviors and health status among MSM using public data from the Blued (Blued International Inc) app.

Methods: A total of 843,745 public postings of 377,610 MSM users located in Guangdong were saved from the Blued app by automatic screen capture. Positive affect, negative affect, sexual behaviors, and health status were measured using the Simplified Chinese Linguistic Inquiry and Word Count. Emotions, including joy, sadness, anger, fear, and disgust, were measured using the Weibo Basic Mood Lexicon. A positive sentiment score and a positive emotion score were also calculated. Univariate and multivariate linear regression models on the basis of a permutation test were used to assess the associations of affective states with sexual behaviors and health status.

Results: A total of 5871 active MSM users and their 477,374 postings were finally selected. Both positive affect and positive emotions (eg, joy) peaked between 7 AM and 9 AM. Negative affect and negative emotions (eg, sadness and disgust) peaked between 2 AM and 4 AM. During that time, 25.1% (97/387) of negative postings were related to health and 13.4% (52/387) of negative postings were related to seeking social support. A multivariate analysis showed that the MSM who were more likely to post sexual behaviors were more likely to express positive affect (beta=0.3107; *P*<.001) and positive emotions (joy: beta=0.027; *P*<.001), as well as negative emotions (sadness: beta=0.0443; *P*<.001 and disgust: beta=0.0256; *P*<.001). They also had a higher positive sentiment score (beta=0.2947; *P*<.001) and a higher positive emotion score (beta=0.1612; *P*<.001). The MSM who were more likely to post their health status were more likely to express negative affect (beta=0.8088; *P*<.001) and negative emotions,

including sadness (beta=0.0705; *P*<.001), anger (beta=0.0058; *P*<.001), fear (beta=0.0052; *P*<.001), and disgust (beta=0.3065; *P*<.001), and less likely to express positive affect (beta=−0.0224; *P*=.02). In addition, they had a lower positive sentiment score (beta=−0.8306; *P*<.001) and a lower positive emotion score (beta=−0.3743; *P*<.001).

Conclusions: The MSM social media community mainly expressed their positive affect in the early morning and negative affect after midnight. Positive affective states were associated with being sexually active, whereas negative affective states were associated with health problems, mostly about mental health. Our finding suggests the potential to deliver different health-related intervention strategies (eg, psychological counseling and safe sex promotion) on a social media app according to the affective states of MSM in real time.

(J Med Internet Res 2020;22(1):e13201) doi:10.2196/13201

KEYWORDS

affect; men who have sex with men; sexual behaviors; health status; social media

Introduction

Background

Affective states, including emotion and sentiment, can be defined as positive or negative evaluations of objects, behaviors, or thoughts [1]. Affective states are critical determinants of health [2,3]. An emotion is a cognition that arouses one or more specific forms of a certain generic type of body reaction [4]. There are 6 basic emotions: joy, sadness, anger, fear, disgust, and surprise. For example, to fear something, such as a snake, is to be cognizing something fearful. The sentiment is a conscious state that develops over time from emotions [5] and is more stable and more permanent than emotions [6]. Sentiment is highly socialized. Suppose a certain object with a complex nature and structure has been repeatedly perceived or thought of by a person in many different contexts on various occasions, these various cognitions of the object will produce a complex dispositional idea of the object, finally leading to strong emotions being felt toward it on many occasions. Anything that excites the dispositional idea of the object will tend to excite all strong emotional dispositions. Summing up all these emotions forms a sentiment about this object. For example, an elderly person feels pride when he or she remembers his or her childhood days. The sentiment includes positive affect (eg, love and honor) and negative affect (eg, hurt and annoyance) [7]. Negative affect and negative emotions are common components of depression, whereas positive affect can facilitate positive health behaviors across the population [8,9]. Therefore, affective states are important components of mental health status to indicate or predict health-related behaviors.

In recent years, with the development of social media apps, an increasing number of studies have suggested that significant links exist between the indicators of emotional well-being or affective states and users' behaviors based on an analysis of user-generated textual data. Golder et al [10] pioneered in using real-time user-generated textual data from Twitter and analyzing the daily changes of positive affect and negative affect, which was the flagship study involving social media data for analyzing emotion-related outcomes. Several following studies have shown the feasibility of using social media data to investigate various health-related issues in the general population, including sleep complaints, depression, anxiety, suicide, and HIV [11-14]. However, similar studies using social media data are rare for the HIV risk population, men who have sex with men (MSM). There was a study that used Grindr, which is a geosocial networking app for MSM in the United States [15]. However, the study used Grindr as the recruitment tool to recruit the MSM population and then conducted a Web-based survey via Grindr but did not derive results using the user-generated data in Grindr [15]. Therefore, there is a lack of studies assessing and monitoring the affective states of MSM by analyzing user-generated data on social media apps.

Blued, like Twitter and Facebook, is a social media site with an app, originating in China, for gay, bisexual men, transgender women, and MSM to communicate with each other to find sexual partners and share information [16,17]. In 2016, Blued reported that it had 27 million users, making it the most widely used gay app in China. There are indications that accessing Blued has become a daily activity for many MSM users: 20% spend at least two hours per day on Blued and 10% post over 200 messages per day, including postings and comments [18]. In addition, owing to its anonymity, MSM can talk freely on Blued about private issues, such as sexual behaviors, health, and HIV status [7,19]. These indicate the potential of using Blued as a valuable data source to assess and monitor affective states among MSM [11,12,20-22].

Using Blued to find male sex partners is related to sexual health and sexually transmitted diseases. There were some questionnaire-based studies on the associations between affective states and sexual behaviors among the general population. It has been found that positive affect and positive emotions (eg, joy) may increase sexual desire that can facilitate sexual behaviors [23], whereas one study has indicated that positive affect does not always translate into behaviors [24]. Similarly, negative affect and negative emotions (eg, sadness or anger) can facilitate or inhibit sexual behaviors [25,26]. For example, anxiety has been found to facilitate genital responses [26], whereas another study suggested that individuals who experienced negative emotions practiced fewer sexual behaviors than those who had positive emotions [25,26]. In the studies of MSM, those with depression or higher negative affect are more likely to practice condomless anal intercourse or have multiple sexual partners [27-29]. Researchers have revealed that affective states may influence the health status among MSM. For example, a depression symptom is closely related to HIV treatment nonadherence, and positive affects predict linkage to HIV care and antiretroviral therapy persistence [30,31]. However, these MSM studies were questionnaire based and

with some opposite results compared with the general population.

Furthermore, mental health greatly depends on affective states [32]. It is found that individuals' inner emotions have been shown to manifest in their choice of words in writing [33,34]. Moreover, studies showed that emotional words used on social media were associated with depression measured by psychological scales [35,36]. The mental health status measured by the questionnaire-based psychological scale needs to be regularly updated to stand the test of time. However, psychological support for depression and anxiety should be timely because of the mental disorder with rapid and unexpected changes. Evidence in mental health research suggests that depression and anxiety are highly prevalent among MSM, for whom depression rates were 40% and 37.6% in the United States and China, respectively [37-39]. Although there are some barriers in using textual data (occurrence of *good* preceded by *not*, recognition of emoticons, etc), it is still a cost-effective way to examine affective states and, therefore, design-specific interventions on psychological support based on social media data [10,40].

Objectives

Therefore, this study was designed to explore affective states, including sentiment and emotions, and their associations with sexual behaviors and health status among Chinese MSM by analyzing user-generated data in Blued. Specifically, this study aimed to understand the following: (1) the description of affective states by displaying the diurnal rhythms of sentiment and emotions and (2) the associations of affective states with sexual behaviors and health status.

Methods

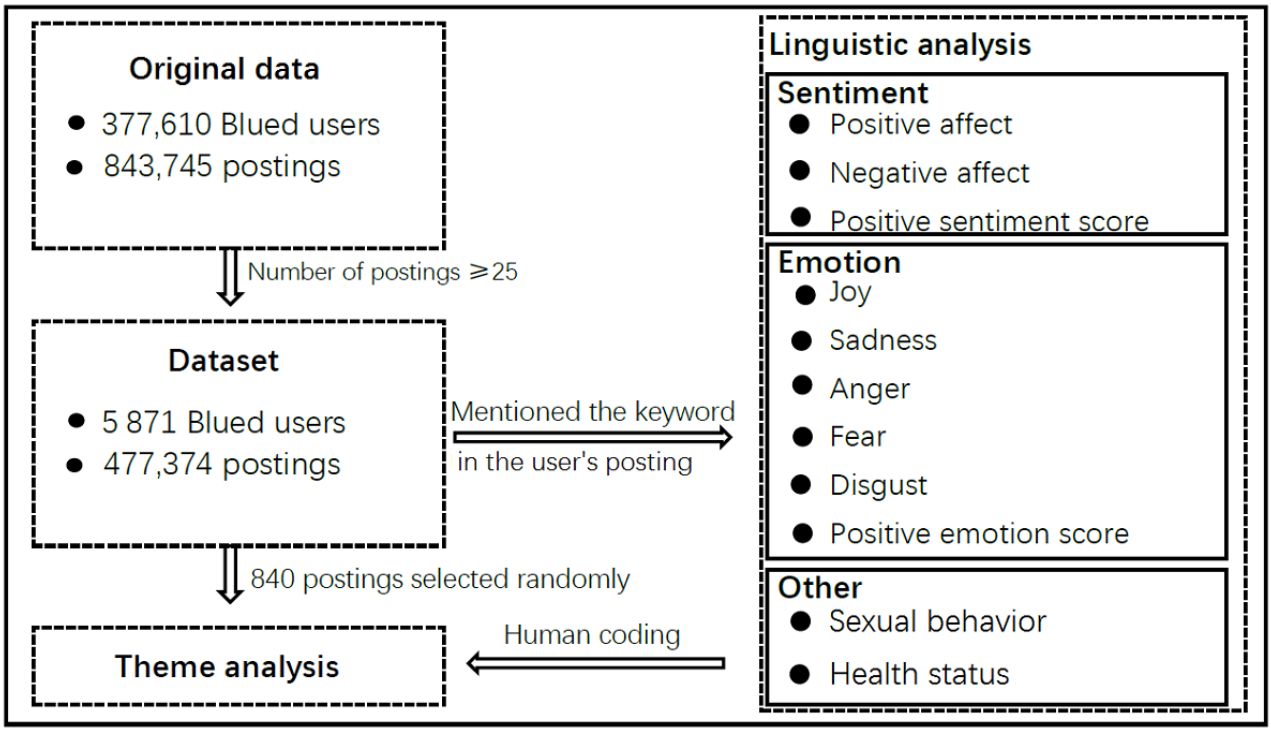
Data Collection

Guangdong is China's most developed province, and it emerged as the earliest province with MSM using the internet in China. Internet-based dating started in Guangdong from the website, GZTZ.ORG, run by the MSM community named as Lingnan, since 1998 [41]. Therefore, the MSM in Guangdong have the longest history of internet-based data and are the largest number of Blued users [18] among all provincial-level divisions in China. Meanwhile, the MSM in China aggregate in the most developed areas [42]. Megacities provide MSM a friendly environment and plenty of opportunities to make friends with each other. Out of 4 megacities in China, 2 megacities, Guangzhou and Shenzhen (GDP per capita is similar to southern European countries), are in the Guangdong province, with a similar Cantonese culture. On the basis of these reasons, we chose the Guangdong province as the target area for our exploratory study. Therefore, from July 2013 to July 2016, 843,745 public postings (N) of 377,610 MSM users (U), including their public background information and postings on Blued, located in Guangdong, were culled from the Blued app through our application programming interface (API). All the data we culled were public information available on Blued, including public background information and postings. The public information data are anonymous. The related participants' identified information has been strictly protected by Blued. In the Blued user agreement, the registered users have agreed that the content posted by users in this app may be copied, reposted, or used by third parties. This study itself does not involve any physical, social, or legal risks to the participants.

To conveniently control various variables involved in our data collection method such as the time interval between 2 simulated clicks, we packaged the three steps of our data collection method into an API. We designed a simulation-based data collection method with three steps: (1) We utilized the android debug bridge (ADB) to simulate the operation of a user using a social network app. Specifically, the ADB can drive a mobile phone to perform a series of actions on the user interface, such as swiping, clicking, or dragging. After these simulated operations, the data (eg, username and profile image) we wanted to collect were displayed on the screen. (2) We used automatic screen capture technology and optical character recognition (OCR) to extract the data displayed on the phone screen. For image data such as profile images and photos, we drew the display zone of each kind of image data according to the user interface of Blued. On the basis of the display zone, we captured the screenshot to obtain the image data we wanted. As for textual data such as username and age, we utilized an open-source OCR tool to extract the text displayed in the screenshot. It is necessary to note that our target texts were of printable style and always displayed in a fixed position in the screenshot. Therefore, the OCR in our method could get high accuracy. (3) We cleaned the data and stored them in a secure device. Owing to the sensitive nature of this work, we took careful steps to protect user privacy and ensure the ethics of the research. All users' data were deidentified in our study. Specifically, we anonymized each user's identity with a random user ID. All data were stored in a secure private server without public access. Only members of our research team had access to log in to this server. Data were transferred using encrypted Secure Sockets Layer connections to ensure that they cannot be intercepted by third parties.

We grouped these data into five databases: user profile database, postings database, comments database, followers database, and followees database. The number of postings is one of the indicators to determine whether a user is an active user. Therefore, users with 25 postings or more were included in the data analysis [40]. Consequently, our sample for this study contained 5871 active users (U), including gay, bisexual men, and transgender women, and 447,374 postings (N; see Figure 1).

Figure 1. Flow chart of the data collection.



Measures

Demographic and Social Network Characteristics

Demographic characteristics, including age, educational level, geolocation, hometown, height, weight, and sex role, were obtained from the user profile database. Social network characteristics, including the number of followers, followees, and chat groups, were also obtained from the user profile database. Age was categorized into *≤25 years old* and *>25 years old* based on the social and developmental psychology literature [11,43]. Educational level was classified as *high school or below*, *above high school*, and *unknown*. Geolocation was categorized into Guangzhou, Shenzhen, Dongguan, and other cities in the Guangdong province based on the distribution of the geolocation of the MSM. Hometown was classified as Guangdong, non-Guangdong, and unknown. Height and weight were transformed into BMI, which was associated with depression as well as sexual activity in the MSM community [44]. Blued users are more likely to search for partners who are in good shape [45]; therefore, BMI is an important indicator for their fitness. BMI was categorized into underweight (BMI<18.5 kg/m^2), normal (18.5 kg/m^2≤BMI<25 kg/m^2), overweight (25 kg/m^2≤BMI<30 kg/m^2), and obese (BMI≥30) on the basis of the guidelines from the National Institutes of Health [44]. In addition, research suggested that those reporting high levels of depressive symptoms were more likely to report both receptive and insertive unprotected anal intercourse [46]. Therefore, sex role was classified into insertive, receptive, and versatile sexual roles and unknown [46]. A *chat group* was defined as an online forum that enables users to conduct instant message-based private conversations with other users. The number of followers, followees, and chat groups, which reflects the social network size or social capital, is associated with depression [47,48]. Owing to the skewed data, the number of followers, followees, and chat groups was log transformed for analysis [49].

Affective States: Sentiment and Emotion

Overview

To measure affective states, we first calculated the probability of words related to affect across all postings for each user in a given hour (P(u,h)). Then, we calculated each user's baseline probability of affect (averaging P(u,h)) across all hours. To measure diurnal mood rhythms, we calculated the grand mean of affect across all users and the relative probability of affect per user. Finally, we calculated the general level of the relative probability of affect across users who were active during hour h. Owing to 7 dependent variables (positive affect, negative affect, and 5 emotions) and many independent variables, there was a possibility of difficulty in exploring the true relationships between them. Thus, to ensure that emotion and sentiment were expressed in one single measure by each posting, the positive emotion score and positive sentiment score were widely applied in different studies [7,50-52]. The calculation of the positive sentiment score and positive emotion score has been described in Step 1. A positive emotion score is different from a positive sentiment score. The true difference is about the concepts of sentiment and emotion. Emotions are preconscious social expressions of feelings and affect [53]. Sentiments are partly social constructs of emotions that develop over time and are enduring [53]. In other words, sentiments have been found to be held for a longer period and are more stable and dispositional than emotions [53]. Therefore, the word list of sentiment is different from the word list of emotion. For example, the word *honor* is a positive affect–related word but is not a joy-related word (Table 1).

Table 1. Summary of the Simplified Chinese Linguistic Inquiry and Word Count and Weibo Basic Mood Lexicon dimensions used for this study and example vocabulary. Example vocabulary in the original Chinese can be found in Multimedia Appendix 1A.

Dimensions	Dictionary	Number of words	Example vocabulary
Sentiment			
Positive affect	SC-LIWC[a]	483	Honor, sweet, happy
Negative affect	SC-LIWC	812	Hurt, agony, nasty
Emotion			
Joy	Weibo-5BML[b]	306	Love, excited, high
Sadness	Weibo-5BML	205	Anxious, alone, tear
Anger	Weibo-5BML	93	Enemy, abuse, roar
Fear	Weibo-5BML	72	Sit on pins and needles, panic, hell
Disgust	Weibo-5BML	142	Wordy, speechless, ridicule
Other			
Sexual-related words	SC-LIWC	117	Sex, condom, kiss
Health-related words	SC-LIWC	375	Infection, insomnia, exercise

[a]SC-LIWC: Simplified Chinese Linguistic Inquiry and Word Count.

[b]Weibo-5BML: Weibo Basic Mood Lexicon.

Step 1

The probability of sentiment and emotions of postings was measured using the Simplified Chinese Linguistic Inquiry and Word Count (SC-LIWC) [54] and Weibo Basic Mood Lexicon (Weibo-5BML) [55]. Owing to the coverage of the word list, using the Chinese version of LIWC (SC-LIWC) was insufficient. Positive affect–, negative affect–, anger-, and sadness-related words are listed in the SC-LIWC. However, joy-, fear-, and disgust-related words are not included in the SC-LIWC. Fortunately, Weibo-5BML, which includes joy, sadness, anger, fear, and disgust, has been developed and can be used as a supplementary tool of LIWC. Therefore, we used SC-LIWC for affect and Weibo-5BML for emotions. A summary of SC-LIWC and Weibo-5BML dimensions' example words for this study is presented in Table 1. This kind of measurement that uses different dictionaries in the same textual data has been applied in Hong's study as well [56]. The validity of LIWC's performance for sentiment has been shown in some studies [54,57-59]. In addition, the validity of Weibo-5BML's performance for emotion has been shown in other studies [55,60,61]. Specifically, Weibo-5BML was validated by comparing the mood time series with fluctuations recorded and labeled by the vital social events and traditional festivals in China [55,60,61]. *Positive affect* and *negative affect* dimensions from SC-LIWC were selected to measure perceived positive and negative affect. Basic emotions, including joy, sadness, anger, fear, and disgust, that correspond to basic emotions identified by Ekman [62] were selected from Weibo-5BML to measure emotional states. The emotion of surprise was excluded because it can be both positive and negative, and it is ambiguous without context [7]. For each user given a posting, we first defined the number of positive emotion–related words (E_u^+) as the difference between the number of 1 positive emotion (joy) related–word and the number of 4 negative emotion (sadness, anger, fear, and disgust)–related words in equation (1). Then, we calculated the positive emotion score using the method described in steps 2 to 5. Similarly, we also estimated the number of positive sentiment–related words (S_u^+) by subtracting the number of positive affect–related words from the number of negative affect–related words in equation (2) and the positive sentiment score [7,50-52]. The positive sentiment score and positive emotion score are the difference between the number of positive sentiments and positive emotions and the number of negative sentiments and negative emotions in each sentence. They indicate that the overall sentiment and emotion in a given posting is rather positive or negative [50].

Step 2

The probability of depression and anxiety is higher for MSM than the general population. Therefore, understanding the frequency and distribution of MSM's emotion-related words across hours gives the clues to find the riskiest period of their status on depression and anxiety, which provides the opportunity for various time-based interventions on mental health targeting MSM. Therefore, sentiment and emotions were measured and displayed across hours, which is similar to one study [40]. For each user in a given hour, we first counted the number of words of postings and the number of words related to positive affect, negative affect, positive sentiment score, joy, sadness, anger, fear, disgust, and positive emotion score in postings. Then, we calculated the following probability (eg, for calculating positive affect):

$$PA(u, h)=||PAWORDS(u, h)||/||WORDS(u, h)|| \quad (3)$$

We used U to index the set of users ($u \in U$) and H to index the set of hours a day ($h \in H$ and $H=\{0,1,2,3\ldots23\}$ (assuming 0–23 for a day)). The measurement of other sentiments or emotions was computed similarly.

Step 3

Regarding variables to assess a user's sentimental and emotional level, for each user, we calculated his baseline probability of affect (averaging P(u,h) across all hours):

$$PA=\Sigma_{hH}\ PA(u,\ h)/||H||\ (4)$$

Note that the baseline probability of affect did not vary from hour to hour and therefore was an indication of the user's average affective state. Consequently, the baseline probability of affect was used as a dependent variable in the multiple linear regression model.

Step 4

To measure diurnal mood rhythms, we then calculated the user's relative probability of affect as defined in equation (5), where the last term is the grand mean across all users over all hours. The relative probability of affect represents the user's deviation from his own baseline probability, which allows us to focus on the user's diurnal mood rhythms by the hour of the day.

$$RPA(u,\ h)=PA(u,\ h)-PA+\Sigma_{(u,\ h)U,H}\ PA(u,\ h)/||UH||\ (5)$$

Step 5

Finally, we calculated the general level of the relative probability affect score as defined in equation (6), where U(h) is the subset of users who were active during hour h [40].

$$RPA(h)=\Sigma_{uU(h)}\ RPA(u,\ h)/||U(h)||\ (6)$$

Sexual Behaviors and Health Status

Sexual behavior–related words (eg, sex, condom, and kiss) and health-related words (eg, infection, insomnia, and exercise) from SC-LIWC were applied to identify the words associated with sexual behaviors and health status in the postings (see Table 1) [58]. The measurement of these 2 dimensions was the same as the measurement of sentiment and emotions. Baseline probabilities of sexual behaviors and health status were calculated as independent variables.

Themes of Postings

Understanding negative affect among MSM is critical for a mental health intervention or crisis interventions. Negative affect peaked between 2 AM and 5 AM, and therefore, 30 postings were selected randomly per time point per day to find out why users had high levels of negative affect. Finally, 840 postings were collected. To identify the themes of postings, especially the negative postings, we scanned all postings to determine the major themes [63]. These postings were categorized into the following 4 themes: (1) expression of emotion and possible cause; (2) expression of health status; (3) expression of sexual behaviors; and (4) expression of emotion. Besides, there was an additional theme categorized for the negative postings: (5) expression of emotion and seeking support.

Statistical Analysis

First, mean and standard deviations were used to describe continuous variables. The frequency was used to describe categorical variables. Next, sentiment and emotion scores displayed a skewed distribution and, therefore, a multiple linear regression model using permutation tests that do not assume normally distributed errors was performed [64]. Univariate and multivariate linear regression models were performed to examine associations between 2 predicted variables—sexual behaviors and health status—and all outcome variables—sentiment (positive affect, negative affect, and positive sentiment score) and emotions (joy, sadness, anger, fear, disgust, and positive emotion score). Variables significant at the $P<.20$ level in the univariate analysis were included and adjusted in the multivariate linear regression model [65]. In total, 840 postings were iteratively coded and sorted into themes by 2 trained assistants separately. To ensure reliability, interrater reliability was measured, and an additional assistant was invited to code the postings that are classed as different themes. The R 3.4.3 version was applied for data analysis [66]. Statistical significance was set at $P<.05$.

Results

Demographic Characteristics

A total of 5871 active MSM users on Blued from the Guangdong province and their 477,374 postings were finally selected. Half of the MSM (3171/5871, 53.99%) were aged 25 years or less, and 35.29% (2072/5871) had an education beyond high school and 13.10% (769/5871) of the MSM had a high school education or below. Most MSM lived in Guangzhou (1999/5871, 34.05%), Shenzhen (1861/5871, 31.70%), or Dongguan (510/5871, 8.69%). Nearly, half of the MSM (2841/5871, 48.39%) identified their hometown as Guangdong, while 36.48% (2142/5871) of the MSM identified their hometown as non-Guangdong. Physically, 75.76% (4448/5871) of the MSM were of normal weight, 16.39% (962/5871) were underweight, 6.64% (390/5871) were overweight, and 1.21% (71/5871) were obese. Sexually, 28.92% (1698/5871) of the MSM self-reported a preference for the insertive sex role, 20.35% (1195/5871) self-reported a preference for the receptive sex role, and 22.55% (1324/5871) self-reported a preference for the versatile sex role (see Table 2).

Affective States: Sentiment

The mean (SD) scores of positive affect, negative affect, and positive sentiment score were 0.014 (0.010), 0.016 (0.012), and −0.001 (0.015), respectively (see Table 3).

Table 2. Sample demographics among men who have sex with men (N=5871).

Characteristics	Value, n (%)
Age (years)	
≤25	3171 (53.99)
>25	2700 (45.99)
Education level	
High school or below	769 (13.10)
Above high school	2074 (35.33)
Missing	3028 (51.58)
Geolocation	
Guangzhou	1999 (34.05)
Shenzhen	1861 (31.70)
Dongguan	510 (8.69)
Other cities in Guangdong	1501 (25.57)
Hometown	
Guangdong	2841 (48.39)
Non-Guangdong	2142 (36.48)
Missing	888 (15.13)
BMI classification	
Underweight	962 (16.39)
Normal	4448 (75.76)
Overweight	390 (6.64)
Fat	71 (1.21)
Sex role	
Receptive	1195 (20.35)
Insertive	1698 (28.92)
Versatile	1324 (22.55)
Missing	1654 (28.17)

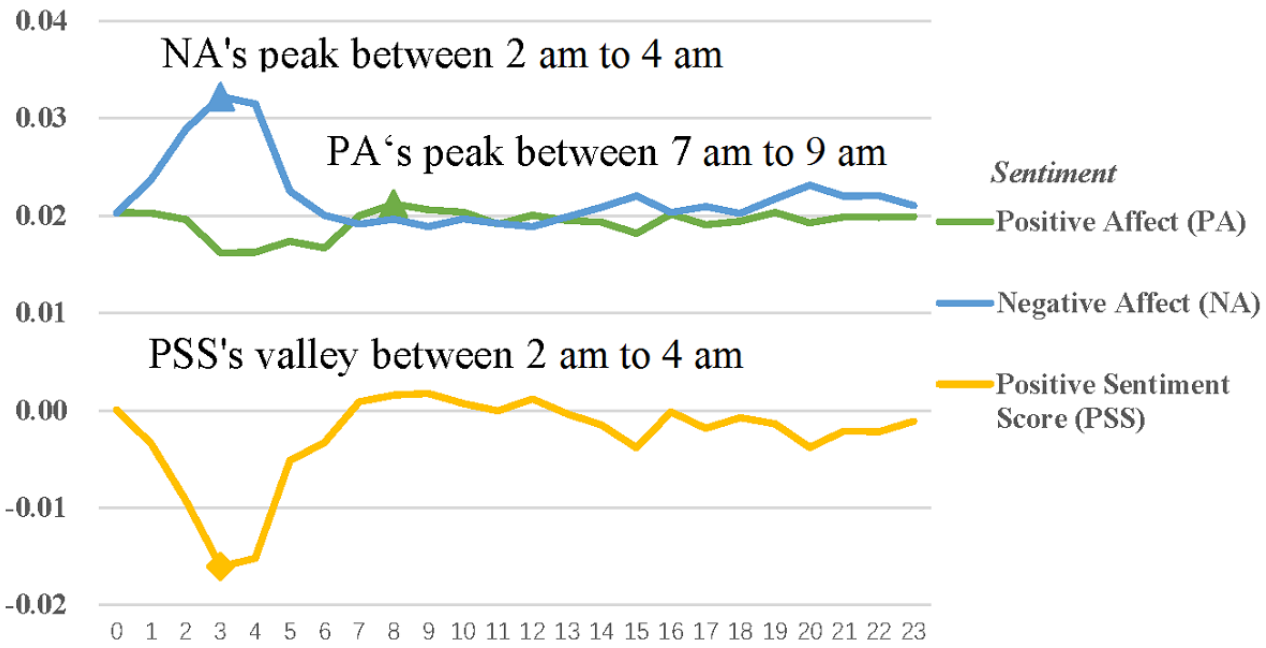
Table 3. Sentiment, emotions, social network variables, sexual behaviors, and health status among men who have sex with men (N=5871).

Characteristics	Value, mean (SD)
Sentiment	
Positive affect	0.014 (0.010)
Negative affect	0.016 (0.012)
Positive sentiment score	−0.001 (0.015)
Emotions	
Joy	0.015 (0.012)
Sadness	0.005 (0.005)
Anger	0.0005 (0.001)
Fear	0.0006 (0.001)
Disgust	0.004 (0.007)
Positive emotion score	0.005 (0.015)
Sexual behavior and health	
Sexual behaviors	0.009 (0.008)
Health status	0.008 (0.008)
Social network variables	
The number of followers (log)	2.341 (0.430)
The number of followees (log)	1.786 (0.889)
The number of chat groups (log)	0.277 (0.331)

Sentiment Diurnal Variation

Positive affect peaked between 7 AM and 9 AM, while negative affect peaked between 2 AM and 4 AM. Specifically, positive affect decreased at noon and increased at 3 AM. Then, positive affect reached the maximum between 7 AM and 9 AM, and next, positive affect showed a relatively stable trend from 9 AM to noon. Negative affect increased at noon and peaked between 2 AM and 4 AM. Then, negative affect decreased between 4 AM and 6 AM. Finally, negative affect also showed a relatively stable trend before noon. The positive sentiment score reached the minimum between 2 AM and 4 AM (Figure 2).

Figure 2. Hourly changes in individual sentiment.



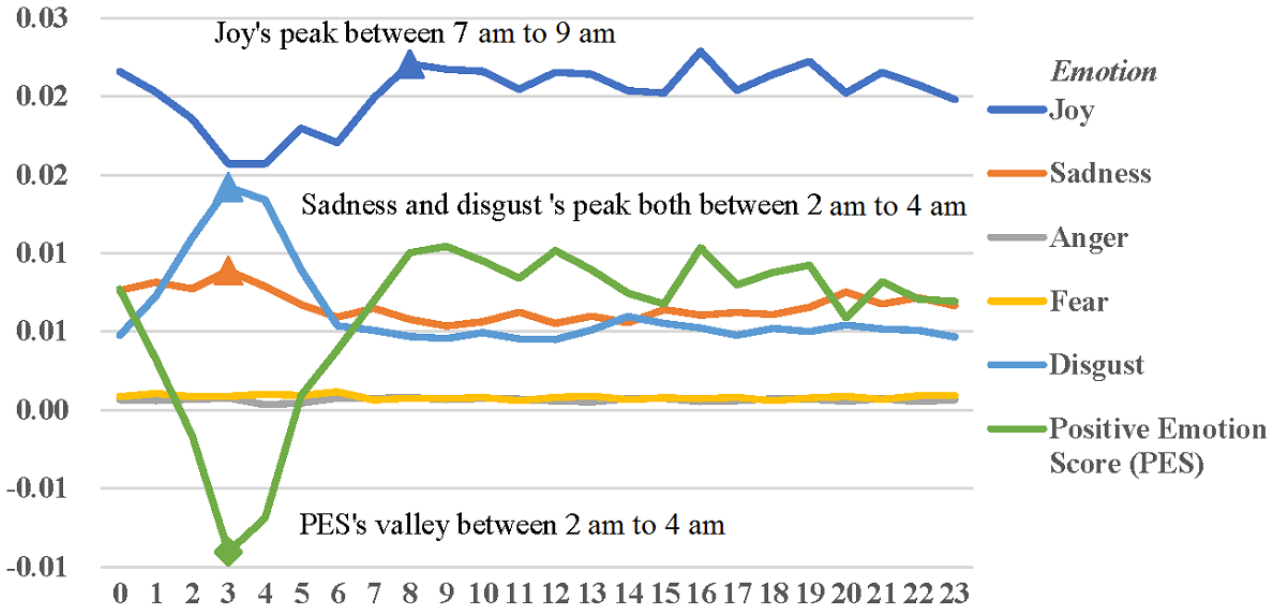
Affective States: Emotion

The mean (SD) scores of joy, sadness, anger, fear, disgust, and positive emotion score were 0.015 (0.012), 0.005 (0.005), 0.0005 (0.001), 0.0006 (0.001), 0.004 (0.007), and 0.005 (0.015), respectively.

Emotion Diurnal Variation

Joy was at a peak between 7 AM and 9 AM, while both sadness and disgust were at a peak between 2 AM and 4 AM. Specifically, joy decreased at noon and increased at 3 AM. Then, joy peaked between 7 AM and 9 AM. Next, joy showed a relatively stable trend but had 2 small peaks at 4 PM and 7 PM. Sadness and disgust both peaked between 2 AM and 4 AM, and both had a relatively stable trend before noon. Anger and fear both showed a relatively stable trend over time. The positive emotion score reached the minimum between 2 AM and 4 AM (Figure 3).

Figure 3. Hourly changes in individual emotions.



Theme of Postings

Negative affect peaked between 2 AM and 5 AM. The proportion of negative, positive, and neutral affect was 46.1% (387/840), 33.8% (284/840), and 20.1% (169/840), respectively, during that time; 79.9% (671/840) of postings contained emotional information. Furthermore, 11.5% (97/840) of postings contained health-related information and 3.1% (26/840) of postings contained sexual behavior–related information.

Among the 387 negative postings, the kappa (agreement) values of the expression of health status, expression of emotion and seeking support, expression of sexual behaviors, expression of emotion and possible cause, and expression of emotion were 0.91 (97%), 0.63 (93%), 0.56 (98%), 0.96 (98%), and 0.92 (99%), respectively. A quarter (97/387, 25.1%) of the negative postings were directly related to health status, among which 86.6% (335/385) were on sleep health or mental health and another 13.4% (52/387) were on physical health. The expression of emotion and seeking support accounted for 13.4% (52/387), and the expression of sexual behaviors accounted for 2.6% (10/387). Of the negative postings, 55.0% (213/387) expressed strong emotion and described possible cause, among which the topics varied, from daily life events (eg, amusement and work; 109/213, 51.2%) and relationship (eg, partnership and friendship; 80/213, 37.5%) to the philosophy of life (24/213, 11.3%); 3.9% (15/387) of postings only expressed strong emotion. Example responses (see Multimedia Appendix 1B-F for the original Chinese text) included the following:

> *Waking up in the middle of the night suddenly and I couldn't fall asleep at night. No one to hug me and feeling lonely.* [Sleep health or mental health.]
>
> *I want to find a boyfriend in Shenzhen, but ugly men couldn't get a boyfriend.* [Expression of emotion and seeking support.]
>
> *Very irritated. I feel I am in a rut and I try my best to control my sexual behaviors.* [Expression of sexual behaviors.]
>
> *Alas, I can't seem to do anything to progress in life. Try to sleep rather than thinking which makes me upset.* [Daily life event.]
>
> *Even though I tried to be strong, I couldn't help but cry.* [Expression of emotion.]

Among the 284 positive postings, the kappa (agreement) values of the expression of health status, expression of sexual behaviors, expression of emotion and possible cause, and expression of emotion were 0.91 (99%), 0.88 (96%), 0.78 (90%), 0.79 (90%), respectively. Only 4.6% (13/284) of the positive postings were related to health status, and a quarter (71/284, 25.0%) of the positive postings were directly related to sexual

behaviors. The expression of emotion accounted for 35.9% (102/284), and the expression of emotion and possible cause accounted for 34.5% (98/284). Example responses (see Multimedia Appendix 1G-J for the original Chinese text) included the following:

> *Gathering sand into a tower~ working, exercising, and doing charity (donation steps) ~ you can also do it.* [Expression of health status; "Gathering sand into a tower" is similar as "a pin a day is a groat a year" in English. For "donation steps," if WeChat, the most popular social network app in China, tracks your steps to be more than 10,000 per day, the WeChat company, Tencent, will donate 2 Renminbi to the charity foundation.]
>
> *They said that being loved is a kind of happiness. The receptive role guyin sex has a feeling like the god in the paradise! What do you think???* [Expression of sexual behaviors.]
>
> *Finally saw the Russia in the snow, satisfied!* [Expression of emotion and possible cause.]
>
> *In a very good mood.* [Expression of emotion.]

Social Network

The mean (SD) scores of followers, followees, and chat groups were 2.341 (0.430), 1.786 (0.889), and 0.277 (0.331) after log transformations, respectively (see Table 3).

Sexual Behaviors and Health Status

The mean (SD) scores of sexual behaviors and health status were 0.009 (0.008) and 0.008 (0.008), respectively. Sexual behaviors peaked between noon and 1 AM. More importantly, there was a similar trend between sexual behaviors, positive affect, and joy (see Figure 4). Health status peaked between 2 AM and 4 AM. In addition, there was a similar trend between health status, negative affect, and disgust (see Figure 5).

Figure 4. Hourly changes in individual sexual behaviors, positive affect, and joy.

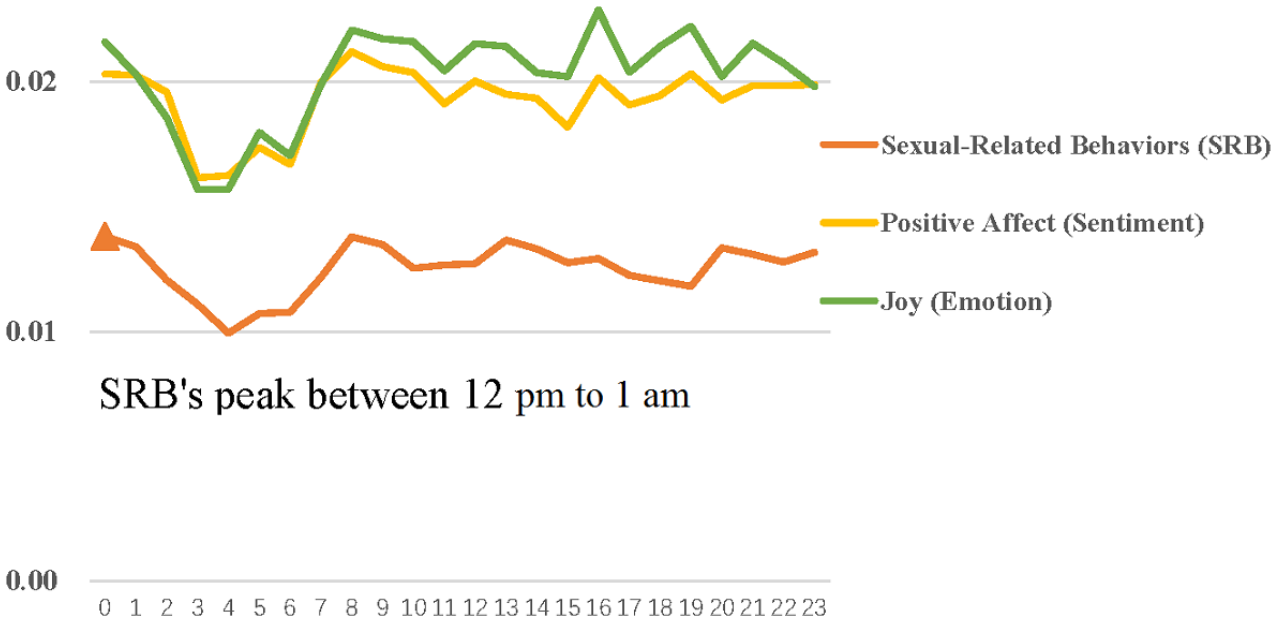
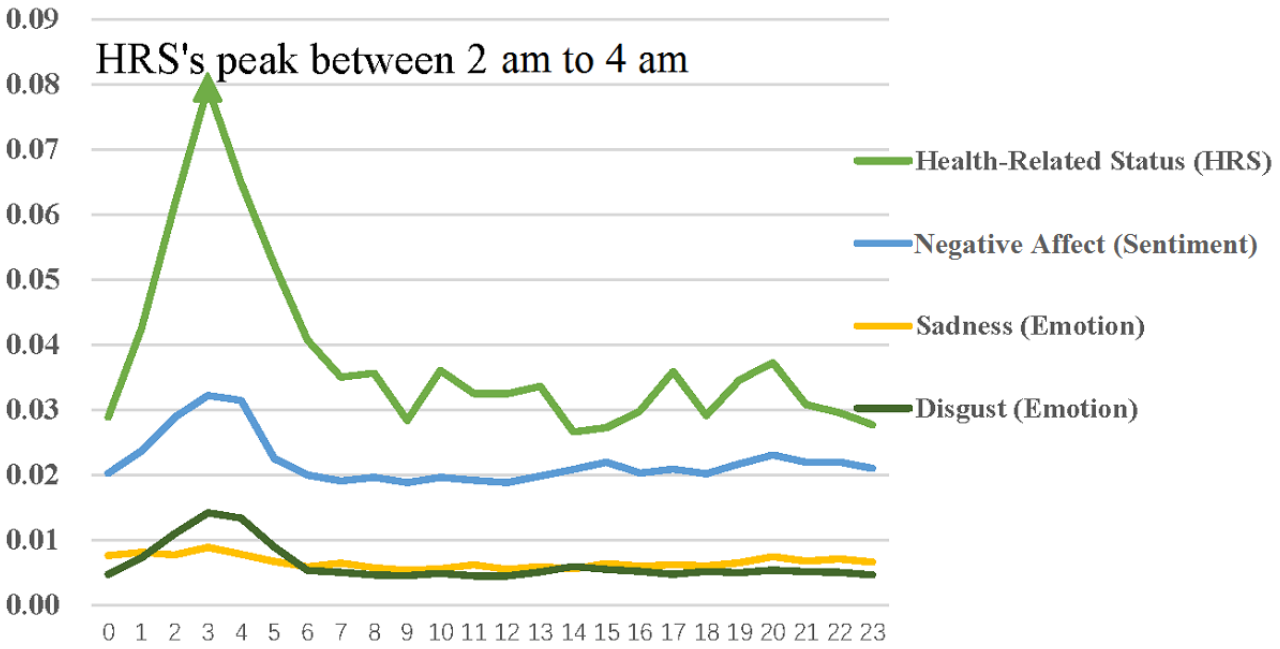


Figure 5. Hourly changes in individual health status, negative affect, sadness, and disgust.



Associations Between Affective States and Demographic Characteristics

Multimedia Appendices 2-4 show the associations between affective states and demographic characteristics. Multivariable models demonstrated that compared with younger MSM (≤25 years), older MSM (>25 years) expressed more negative affect (beta=0.0004; *P*<.001), more negative emotions (anger: beta=0.00003; *P*<.001 and fear: beta=0.00003; *P*=.04), less positive affect (beta=−0.0004; *P*<.001), and a less negative emotion (disgust: beta=−0.0001; *P*=.03) and had a lower positive sentiment score (beta=−.0009; *P*<.001) and a lower positive emotion score (beta=−0.0002; *P*=.03). Compared with the MSM with a high school education or below, the MSM with an education beyond high school expressed more negative emotions (sadness: beta=0.0004; *P*=.02 and fear: beta=0.0001; *P*<.001) and had a lower positive sentiment score (beta=−0.0009; *P*=.03). Compared with the MSM who lived in Guangzhou, those who lived in Shenzhen expressed less negative emotions (sadness: beta=−0.0004; *P*<.001 and anger: beta=−0.00001; *P*=.049); the MSM who lived in Dongguan expressed less positive affect (beta=−0.0004; *P*<.001) and a less positive emotion (joy: beta=−0.0007; *P*<.001); the MSM who lived in other cities in Guangdong expressed more negative affect (beta=0.0006; *P*<.001) and a more positive emotion (joy: beta=0.001; *P*<.001). Compared with the MSM whose hometowns were in Guangdong, non-Guangdong MSM expressed a more positive emotion (joy: beta=0.0003; *P*=.02), a more negative emotion (fear: beta=0.00002; *P*=.005), less negative affect (beta=−0.0003; *P*<.001), and a less negative emotion (anger: beta=−0.00005; *P*<.001) and had a higher positive sentiment score (beta=0.0007; *P*<.001) and a higher positive emotion score (beta=0.0006; *P*=.02). Compared with the MSM with normal weight, underweight MSM expressed less positive affect (beta=−0.0004; *P*=.02) and less negative emotions (anger: beta=−0.00007; *P*<.001 and fear: beta=−0.00006; *P*=.03). Compared with the MSM with normal weight, overweight MSM expressed more negative affect (beta=0.0008; *P*<.001) and a more negative emotion (sadness: beta=0.0004; *P*<.001) and had a lower positive sentiment score (beta=−0.0008; *P*=.04). Obese MSM expressed less positive affect (beta=−0.0004; *P*=.04) and more negative affect (beta=0.0006; *P*=.03) and had a lower positive sentiment score (beta=−0.0009; *P*=.03). Compared with the MSM who prefer receptive anal intercourse, those preferring insertive anal intercourse expressed less negative affect (beta=−0.0003; *P*<.001), whereas those preferring versatile anal intercourse expressed more negative affect (beta=0.0001; *P*=.04) and a less negative emotion (fear: beta=−0.000002; *P*=.03).

Associations Between Affective States and Social Networks

In multivariable models (Multimedia Appendices 2-4), the MSM with more chat groups expressed more positive affect (beta=0.0008; *P*<.001), a more positive emotion (joy: beta=0.0022; *P*<.001), a less negative emotion (sadness: beta=−0.0003; *P*=.03), a higher positive sentiment score (beta=0.0011; *P*<.001), and a higher positive emotion score (beta=0.0024; *P*<.001). The MSM with more followees expressed more negative emotions, including sadness (beta=0.0002; *P*<.001) and disgust (beta=0.0003; *P*=.002). The MSM with more followers expressed more positive affect (beta=0.002; *P*<.001), a more positive emotion (joy: beta=0.003; *P*<.001), a less negative emotion (sadness: beta=−0.0001; *P*=.04), a more negative emotion (disgust: beta=0.003; *P*=.02), a higher positive sentiment score (beta=0.002; *P*<.001), and a higher positive emotion score (beta=0.0026; *P*<.001).

Associations Between Affective States and Sexual Behaviors

In multivariable models (Multimedia Appendices 2-4), the MSM who were more likely to post sexual behaviors not only expressed more positive affect (beta=0.3107; *P*<.001) and a positive emotion (joy: beta=0.027; *P*<.001) but also expressed more negative emotions, including sadness (beta=0.0443; *P*<.001) and disgust (beta=0.0256; *P*<.001). More importantly, the MSM who were more likely to post sexual behaviors had a higher positive sentiment score (beta=0.2947; *P*<.001) and a higher positive emotion score (beta=0.1612; *P*<.001).

Associations Between Affective States and Health Status

In multivariable models (Multimedia Appendices 2-4), the MSM who were more likely to post their health status not only expressed less positive affect (beta=−0.0224; *P*=.02) but also expressed more negative affect (beta=0.8088; *P*<.001) and negative emotions, including sadness (beta=0.0705; *P*<.001), anger (beta=0.0058; *P*<.001), fear (beta=0.0052; *P*<.001), and disgust (beta=0.3065; *P*<.001). Moreover, the MSM who were more likely to post their health status had a lower positive sentiment score (beta=−0.8306; *P*<.001) and a lower positive emotion score (beta=−0.3743; *P*<.001).

Discussion

Principal Findings

This is one of the first studies to assess the affective states of MSM using social media data. In addition, this study is the first to explore affective states as the factor associated with sexual behaviors and health status among the MSM population.

This study showed that the MSM population was more likely to express negative affect and negative emotions (sadness and disgust) between 2 AM and 4 AM. In addition, a quarter of the negative postings were directly related to MSM's health and about one-eighth reported that MSM needed social support during that sensitive period. Owing to the high prevalence of depression and strong links between the emotional words used and clinical depression among MSM, it is essential to implement interventions (eg, providing Web-based psychological counseling or tailored risk reduction reminders) based on this app [37-39]. Few studies have assessed the affective states of MSM based on social media, but some studies have shown that diurnal mood swings reflect endogenous circadian rhythms interacting with the duration of prior sleep or wakefulness. A previous study showed that positive affect rose quickly from 9 AM to noon and remained steady until 9 PM, after which it fell sharply [67]. In other words, positive affect peaked between midnight and 9 PM [67]. Another study found that positive emotions (happy, warm, and enjoying) had 2 peaks at noon and in the evening [68]. A time-series study found that positive affect had 2 peaks in the afternoon and evening [69]. Another study found that positive affect had 2 peaks: relatively early in the morning and again near midnight [40]. In this study, positive affect and positive emotions (eg, joy) peaked in the morning and relatively plateaued from morning to evening, which is different from other studies [40,67-70]. As for negative affect and negative emotions, studies found that negative affect peaked in the afternoon [69] and evening [40,69]. Negative emotions (eg, depressed or blue, hassled, criticized, worry, and angry) had 2 peaks at midmorning and midafternoon [68]. Besides, several studies have found that negative affect is not subject to diurnal variation [67,70]. This study found that negative affect and negative emotions (eg, sadness and disgust) peaked in the evening, which was similar to a previous study [40]. However, negative affect and negative emotions also relatively plateaued from morning to evening. There are several reasons for the diurnal variation in mood among MSM. First, for some people, the symptoms of depression may be worse at night, leading to difficulty in getting to sleep and to a feeling of isolation and hopelessness. Second, owing to prejudice, stigma, and social pressures for MSM, they may be more likely to use gay apps (eg, Blued) to express their mood in the evening. Moreover, MSM may tend to avoid discussing their health, needs, and mood in real life [71]. However, Blued provides an anonymous environment for MSM to exchange opinions and share information, which possibly explains why these men tended to talk more openly about health-related topics, express affect, and seek help [19]. These findings highlight the unmet emotional requests of MSM.

The study found that sexual behaviors were associated with positive affect, positive emotions, and negative emotions. More importantly, the MSM who were more likely to post sexual behaviors had a higher positive sentiment score and a higher positive emotion score. Previous studies found that positive affect facilitated sexual behaviors [23,72]. It is generally acknowledged that positive affect can increase sexual arousal and sexual desire and therefore facilitate sexual behavior [23]. In contrast, the MSM with negative emotions may use sex as a mood regulator and may practice more condomless anal sex with casual partners [27,73-75]. One possible explanation for this finding is that health-compromising behaviors (eg, unprotected sex) may be used as coping mechanisms to manage the effects of negative emotions. Owing to the high prevalence of HIV among MSM, practicing riskier sexual behaviors, in turn, may lead to a higher psychological burden [76]. In general, sexual behaviors seemed to be more associated with positive affect and positive emotions in this study.

Finally, the MSM who are more likely to post their health status may express more negative affect and negative emotions. Positive affect facilitates positive health behavior, leading to favorable health outcomes such as fewer symptoms and less pain [9]. Nevertheless, negative affect and negative emotions usually coexist with illness and may influence one's functional status and health-related quality of life, meaning that measuring negative affect and negative emotions may provide a valuable means for understanding the health of MSM [77].

Limitations

It is important to acknowledge the limitations of this study. First, we only used data from the Guangdong province, which may affect generalization owing to the cultural differences in different geographic locations. In addition, the generalizability of this study was also affected by the characteristics of the study participants. Compared with a community sample of Chinese

MSM, the internet sample was significantly younger and more educated [19]. Moreover, compared with other studies using millions of users' messages, the sample size is not big enough to measure affective states precisely [40]. Second, although keyword research was used to measure sexual behaviors and health status, social media data may not be fully representative of a user's actual behaviors, indicating that a combination of social media data and survey research is essential to understanding the association between sexual behaviors and health status and affective states.

Conclusions

The MSM social media community mainly expressed their positive affect in the early morning and negative affect after midnight. Positive affective states were associated with being sexually active, whereas negative affective states were associated with health problems, mostly about mental health. The findings of this study suggest different health-related intervention strategies for MSM app developers and users. MSM app developers can consider switching the banner of the app's home page or prioritizing the posts or users over time with algorithms according to the affective states of MSM. For example, the reminders of safe sex on the home page can be displayed in the early morning, while heartwarming slogans or psychological counseling links can be exhibited in the wee hours. Furthermore, social media–based psychological assistance can narrow down to focus more on users who stably expressed negative affects, whereas the sexual risk reduction interventions can focus on those who stably expressed positive affects. For example, the content with sexual health issues could be prioritized for the users with positive sentiment states, whereas the content with encouragement could be prioritized for the users with negative sentiment states. This study shows the potential of using social media to support MSM with health issues. Future online-offline integrating surveys are warranted to confirm the outcomes of the health (eg, depression, infections, etc) and health-related behaviors (eg, sexual behaviors) of MSM for providing precision interventions to those who are most in need.

Acknowledgments

We are grateful to the participants of this study. This project was funded by the National Natural Science Foundation of China (grant number #91546203, #71974212, #71774178, and #81803334), Science and Technology Program of Guangzhou, China (grant number #201607010332) and A Major Infectious Disease Prevention and Control of the National Science and Technology Major Project (grant number #2018ZX10715004).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quoted vocabulary and responses in the original Chinese text.
[DOCX File , 17 KB - jmir_v22i1e13201_app1.docx]

Multimedia Appendix 2
Univariate and multivariate analysis of sentiment, sexual behaviors, and health status.
[DOCX File , 86 KB - jmir_v22i1e13201_app2.docx]

Multimedia Appendix 3
Univariate and multivariate analysis of emotions (joy, sadness, and disgust), sexual behaviors, and health status.
[DOCX File , 86 KB - jmir_v22i1e13201_app3.docx]

Multimedia Appendix 4
Univariate and multivariate analysis of emotions (anger, fear, and positive emotion score), sexual behaviors, and health status.
[DOCX File , 86 KB - jmir_v22i1e13201_app4.docx]

References

1. Thoits PA. The sociology of emotions. Annu Rev Sociol 1989;15(1):317-342 [FREE Full text] [doi: 10.1146/annurev.so.15.080189.001533]
2. Desmet P. Designing emotions. Des J 2002;6(2):60-62 [FREE Full text]
3. Khosrow-Pour DB. Encyclopedia of Information Science and Technology. Fourth Edition. Pennsylvania, United States: IGI Global; 2018.
4. Broad CD. Emotion and sentiment. J Aesthet Art Crit 1954;13(2):203-214. [doi: 10.2307/425913]
5. Stets JE. Emotions and sentiments. In: DeLamater J, editor. Handbook of Social Psychology. Boston, MA: Springer; 2006:309-335.
6. Ben-Ze'ev A. The Subtlety of Emotions. Cambridge: The MIT Press; 2000.

7. Volkova S, Bachrach Y. On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. Cyberpsychol Behav Soc Netw 2015 Dec;18(12):726-736. [doi: 10.1089/cyber.2014.0609] [Medline: 26652673]
8. Anas M, Akhouri D. Positive and Negative Affect in Depressed and Normal Adults. J Indian Health Psychol 2013;8(1):61-68 [FREE Full text]
9. van Cappellen P, Rice EL, Catalino LI, Fredrickson BL. Positive affective processes underlie positive health behaviour change. Psychol Health 2018 Jan;33(1):77-97 [FREE Full text] [doi: 10.1080/08870446.2017.1320798] [Medline: 28498722]
10. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]
11. Tian X, Yu G, He F. An analysis of sleep complaints on Sina Weibo. Comput Hum Behav 2016;62:230-235 [FREE Full text] [doi: 10.1016/j.chb.2016.04.014]
12. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
13. Tian X, He F, Batterham P, Wang Z, Yu G. An analysis of anxiety-related postings on Sina Weibo. Int J Environ Res Public Health 2017 Jul 13;14(7):pii: E775 [FREE Full text] [doi: 10.3390/ijerph14070775] [Medline: 28703773]
14. Jung H, Park H, Song T. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259 [FREE Full text] [doi: 10.2196/jmir.7452] [Medline: 28739560]
15. Gibbs JJ, Rice E. The social context of depression symptomology in sexual minority male youth: determinants of depression in a sample of Grindr users. J Homosex 2016;63(2):278-299. [doi: 10.1080/00918369.2015.1083773] [Medline: 26295497]
16. Hong H, Xu J, McGoogan J, Dong H, Xu G, Wu Z. Relationship between the use of gay mobile phone applications and HIV infection among men who have sex with men in Ningbo, China: a cross-sectional study. Int J STD AIDS 2018 Apr;29(5):491-497. [doi: 10.1177/0956462417738468] [Medline: 29099328]
17. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
18. 199IT. 2016. Blued: 2015 Big Data White Paper for Chinese Gay Community URL: http://www.199it.com/archives/428104.html [accessed 2019-06-15]
19. Han J, Tian X, Yu G, He F. Disclosure pattern of self-labeled people living with HIV/AIDS on Chinese social networking site: an exploratory study. Cyberpsychol Behav Soc Netw 2016 Aug;19(8):516-523. [doi: 10.1089/cyber.2016.0133] [Medline: 27494331]
20. Liu P, Tov W, Kosinski M, Stillwell DJ, Qiu L. Do Facebook status updates reflect subjective well-being? Cyberpsychol Behav Soc Netw 2015 Jul;18(7):373-379. [doi: 10.1089/cyber.2015.0022] [Medline: 26167835]
21. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. J Med Internet Res 2017 Jul 10;19(7):e243 [FREE Full text] [doi: 10.2196/jmir.7276] [Medline: 28694239]
22. Yeo TED, Chu TH. Sharing 'sex secrets' on Facebook: a content analysis of youth peer communication and advice exchange on social media about sexual health and intimate relations. J Health Commun 2017 Sep;22(9):753-762. [doi: 10.1080/10810730.2017.1347217] [Medline: 28796578]
23. Wood IB, Varela PL, Bollen J, Rocha LM, Gonçalves-Sá J. Human sexual cycles are driven by culture and match collective moods. Sci Rep 2017 Dec 21;7(1):17973 [FREE Full text] [doi: 10.1038/s41598-017-18262-5] [Medline: 29269945]
24. Janssen E, Macapagal KR, Mustanski B. Individual differences in the effects of mood on sexuality: the revised Mood and Sexuality Questionnaire (MSQ-R). J Sex Res 2013;50(7):676-687 [FREE Full text] [doi: 10.1080/00224499.2012.684251] [Medline: 22963331]
25. von Sadovszky V, Vahey DC, McKinney K, Keller ML. Emotions involved in college students' sexual encounters. West J Nurs Res 2006 Nov;28(7):864-79; discussion 880. [doi: 10.1177/0193945906286805] [Medline: 17056777]
26. Hamilton LD, Meston CM. The role of salivary cortisol and DHEA-S in response to sexual, humorous, and anxiety-inducing stimuli. Horm Behav 2011 May;59(5):765-771 [FREE Full text] [doi: 10.1016/j.yhbeh.2010.12.011] [Medline: 21195074]
27. Babowitch JD, Mitzel LD, Vanable PA, Sweeney SM. Depressive symptoms and condomless sex among men who have sex with men living with HIV: a curvilinear association. Arch Sex Behav 2018 Oct;47(7):2035-2040. [doi: 10.1007/s10508-017-1105-3] [Medline: 29110113]
28. Derefinko KJ, Peters JR, Eisenlohr-Moul TA, Walsh EC, Adams ZW, Lynam DR. Relations between trait impulsivity, behavioral impulsivity, physiological arousal, and risky sexual behavior among young men. Arch Sex Behav 2014 Aug;43(6):1149-1158 [FREE Full text] [doi: 10.1007/s10508-014-0327-x] [Medline: 24958252]
29. Mustanski B, Garofalo R, Herrick A, Donenberg G. Psychosocial health problems increase risk for HIV among urban young men who have sex with men: preliminary evidence of a syndemic in need of attention. Ann Behav Med 2007 Aug;34(1):37-45 [FREE Full text] [doi: 10.1007/bf02879919] [Medline: 17688395]
30. Gonzalez JS, Batchelder AW, Psaros C, Safren SA. Depression and HIV/AIDS treatment nonadherence: a review and meta-analysis. J Acquir Immune Defic Syndr 2011 Oct 1;58(2):181-187 [FREE Full text] [doi: 10.1097/QAI.0b013e31822d490a] [Medline: 21857529]

31. Moskowitz JT. Positive affect predicts lower risk of AIDS mortality. Psychosom Med 2003;65(4):620-626. [doi: 10.1097/01.psy.0000073873.74829.23] [Medline: 12883113]
32. Gross JJ, Uusberg H, Uusberg A. Mental illness and well-being: an affect regulation perspective. World Psychiatry 2019 Jun;18(2):130-139 [FREE Full text] [doi: 10.1002/wps.20618] [Medline: 31059626]
33. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
34. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
35. Settanni M, Marengo D. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Front Psychol 2015;6:1045 [FREE Full text] [doi: 10.3389/fpsyg.2015.01045] [Medline: 26257692]
36. De Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci'13; May 2–4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
37. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 2008 Aug 18;8:70 [FREE Full text] [doi: 10.1186/1471-244X-8-70] [Medline: 18706118]
38. Dyer TP, Regan R, Wilton L, Harawa NT, Ou SS, Wang L, et al. Differences in substance use, psychosocial characteristics and HIV-related sexual risk behavior between Black men who have sex with men only (BMSMO) and Black men who have sex with men and women (BMSMW) in six US cities. J Urban Health 2013 Dec;90(6):1181-1193 [FREE Full text] [doi: 10.1007/s11524-013-9811-1] [Medline: 23897039]
39. Ibragimov U, Harnisch JA, Nehl EJ, He N, Zheng T, Ding Y, et al. Estimating self-reported sex practices, drug use, depression, and intimate partner violence among MSM in China: a comparison of three recruitment methods. AIDS Care 2017 Jan;29(1):125-131. [doi: 10.1080/09540121.2016.1201191] [Medline: 27367038]
40. Gerritsen C, van Breda WR. Simulation-Based Prediction and Analysis of Collective Emotional States. In: Proceedings of the International Conference on Social Computing and Social Media. 2015 Presented at: SCSM'15; August 2-7, 2015; Los Angeles p. 118-126 URL: https://doi.org/10.1007/978-3-319-20367-6_13 [doi: 10.1007/978-3-319-20367-6_13]
41. Hvistendahl M. Public health. China partners with gay groups on HIV screening. Science 2013 Jan 4;339(6115):17-18. [doi: 10.1126/science.339.6115.17] [Medline: 23288517]
42. Qin Q, Guo W, Tang W, Mahapatra T, Wang L, Zhang N, et al. Spatial analysis of the human immunodeficiency virus epidemic among men who have sex with men in China, 2006-2015. Clin Infect Dis 2017 Apr 1;64(7):956-963 [FREE Full text] [doi: 10.1093/cid/cix031] [Medline: 28362948]
43. Stordal E, Mykletun A, Dahl AA. The association between age and depression in the general population: a multivariate examination. Acta Psychiatr Scand 2003 Feb;107(2):132-141. [doi: 10.1034/j.1600-0447.2003.02056.x] [Medline: 12534439]
44. Blashill AJ, Mayer KH, Crane HM, Baker JS, Willig JH, Willig AL, et al. Body mass index, depression, and condom use among HIV-infected men who have sex with men: a longitudinal moderation analysis. Arch Sex Behav 2014 May;43(4):729-734 [FREE Full text] [doi: 10.1007/s10508-013-0155-4] [Medline: 23979782]
45. Meanley S, Hickok A, Johns MM, Pingel ES, Bauermeister JA. Body mass index, body esteem, and unprotected receptive anal intercourse among young men who have sex with men who seek partners online. Arch Sex Behav 2014 May;43(4):735-744 [FREE Full text] [doi: 10.1007/s10508-013-0159-0] [Medline: 23979783]
46. Fendrich M, Avci O, Johnson TP, Mackesy-Amiti ME. Depression, substance use and HIV risk in a probability sample of men who have sex with men. Addict Behav 2013 Mar;38(3):1715-1718 [FREE Full text] [doi: 10.1016/j.addbeh.2012.09.005] [Medline: 23254224]
47. Lee E, Kim YJ, Ahn J. How do people use Facebook features to manage social capital? Comput Hum Behav 2014;36:440-445. [doi: 10.1016/j.chb.2014.04.007]
48. Hatzenbuehler ML, McLaughlin KA, Xuan Z. Social networks and risk for depressive symptoms in a national sample of sexual minority youth. Soc Sci Med 2012 Oct;75(7):1184-1191 [FREE Full text] [doi: 10.1016/j.socscimed.2012.05.030] [Medline: 22771037]
49. Feng C, Wang H, Lu N, Chen T, He H, Lu Y, et al. Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry 2014 Apr;26(2):105-109 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2014.02.009] [Medline: 25092958]
50. Ferrara E, Yang Z. Measuring emotional contagion in social media. PLoS One 2015;10(11):e0142390 [FREE Full text] [doi: 10.1371/journal.pone.0142390] [Medline: 26544688]
51. Ferrara E, Yang Z, Cattuto C. Quantifying the effect of sentiment on information diffusion in social media. PeerJ Comput Sci 2015;1(51). [doi: 10.7717/peerj-cs.26]
52. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. J Manag Inf Syst 2014;29(4):217-248. [doi: 10.2753/MIS0742-1222290408]
53. Munezero M, Montero CS, Sutinen E, Pajunen J. Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. IEEE Trans Affective Comput 2014;5(2):101-111. [doi: 10.1109/TAFFC.2014.2317187]

54. Zhao N, Jiao D, Bai S, Zhu T. Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. PLoS One 2016;11(6):e0157947 [FREE Full text] [doi: 10.1371/journal.pone.0157947] [Medline: 27322382]
55. Dong YH, Chen H, Qian WN, Zhou AY. Micro-blog social moods and Chinese stock market: the influence of emotional valence and arousal on Shanghai Composite Index volume. Int J Embed Syst 2015;7(2):148. [doi: 10.1504/IJES.2015.069987]
56. Hong K, Nenkova A, March ME, Parker AP, Verma R, Kohler CG. Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. Psychiatry Res 2015 Jan 30;225(1-2):40-49. [doi: 10.1016/j.psychres.2014.10.002] [Medline: 25480546]
57. Wang J, Gan SQ, Zhao N, Liu T, Zhu T. Chinese mood variation analysis based on Sina Weibo. J University Chinese Acad Sci 2016;33(6):815-824. [doi: 10.7523/j.issn.2095-6134.2016.06.014]
58. Gao R, Hao B, Li H, Zhu T, Gao Y. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. In: Proceedings of the International Conference on Brain and Health Informatics. 2013 Presented at: BHI'13; October 29-31, 2013; Maebashi, Japan p. 359-368. [doi: 10.1007/978-3-319-02753-1_36]
59. Zhang XY. LIWC: A Text Analysis Tools Based on Word Count. Journal of Southwest University for Nationalities (Humanities and Social Science) 2015:101-104. [doi: 10.1007/978-3-319-20702-5_10]
60. Dong YH, Chen G, Lai KQ, Yue GA. Weibo Social moods measurement and validation. J Psychol Sci 2015(5):1141-1146 [FREE Full text]
61. Dong Y, Chen H, Tang X, Qian W, Zhou A. Collective Emotional Reaction to Societal Risks in China. In: Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics. 2015 Presented at: IEEE SMC'15; October 9-12, 2015; Kowloon, China. [doi: 10.1109/SMC.2015.108]
62. Ekman P. An argument for basic emotions. Cogn Emot 1992;6(3-4):169-200. [doi: 10.1080/02699939208411068]
63. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2003;15(1):85-109. [doi: 10.1177/1525822x02239569]
64. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage 2014 May 15;92:381-397 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.01.060] [Medline: 24530839]
65. Smith LW, Guy R, Degenhardt L, Yeung A, Rissel C, Richters J, et al. Meeting sexual partners through internet sites and smartphone apps in Australia: National Representative Study. J Med Internet Res 2018 Dec 18;20(12):e10683 [FREE Full text] [doi: 10.2196/10683] [Medline: 30563809]
66. R: The R Project for Statistical Computing. 2019. URL: https://www.r-project.org/ [accessed 2019-10-25]
67. Clark LA, Watson D, Leeka J. Diurnal variation in the positive affects. Motiv Emot 1989;13(3):205-234. [doi: 10.1007/BF00995536]
68. Stone AA, Schwartz JE, Schkade D, Schwarz N, Krueger A, Kahneman D. A population approach to the study of emotion: diurnal rhythms of a working day examined with the Day Reconstruction Method. Emotion 2006 Feb;6(1):139-149. [doi: 10.1037/1528-3542.6.1.139] [Medline: 16637757]
69. Vittengl JR, Holt CS. A time-series diary study of mood and social interaction. Motiv Emot 1998;22(3):255-275. [doi: 10.1023/A:1022388123550]
70. Hasler BP, Mehl MR, Bootzin RR, Vazire S. Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. J Res Pers 2008;42(6):1537-1546. [doi: 10.1016/j.jrp.2008.07.012]
71. Tang W, Mahapatra T, Tucker JD, Zhao J, Detels R. Prejudice, stigma and social discrimination: perceived barriers in HIV prevention and control among men who have sex with men in China. J Clin Case Rep 2014;4(11):448. [doi: 10.4172/2165-7920.1000448]
72. Shrier LA, Shih MC, Beardslee WR. Affect and sexual behavior in adolescents: a review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. Pediatrics 2005 May;115(5):e573-e581 [FREE Full text] [doi: 10.1542/peds.2004-2073] [Medline: 15867022]
73. Bancroft J, Janssen E, Strong D, Vukadinovic Z. The relation between mood and sexuality in gay men. Arch Sex Behav 2003 Jun;32(3):231-242. [doi: 10.1023/a:1023461500810] [Medline: 12807295]
74. Millar BM, Starks TJ, Grov C, Parsons JT. Sexual Risk-Taking in HIV-Negative Gay and Bisexual Men Increases with Depression: Results from a US National Study. AIDS Behav 2017 Jun;21(6):1665-1675 [FREE Full text] [doi: 10.1007/s10461-016-1507-6] [Medline: 27475943]
75. Bousman C, Cherner M, Ake C, Letendre S, Atkinson J, Patterson T, HNRC Group. Negative mood and sexual behavior among non-monogamous men who have sex with men in the context of methamphetamine and HIV. J Affect Disord 2009 Dec;119(1-3):84-91 [FREE Full text] [doi: 10.1016/j.jad.2009.04.006] [Medline: 19419773]
76. Walsh JL, Senn TE, Carey MP. Longitudinal associations between health behaviors and mental health in low-income adults. Transl Behav Med 2013 Mar;3(1):104-113 [FREE Full text] [doi: 10.1007/s13142-012-0189-5] [Medline: 23997836]
77. Hu J, Gruber KJ. Positive and negative affect and health functioning indicators among older adults with chronic illnesses. Issues Ment Health Nurs 2008 Aug;29(8):895-911. [doi: 10.1080/01612840802182938] [Medline: 18649214]

Abbreviations

ADB: android debug bridge

API: application programming interface
MSM: men who have sex with men
OCR: optical character recognition
SC-LIWC: Simplified Chinese Linguistic Inquiry and Word Count
Weibo-5BML: Weibo Basic Mood Lexicon

Edited by G Eysenbach; submitted 20.12.18; peer-reviewed by S Bassett, H Tsai; comments to author 28.03.19; revised version received 24.06.19; accepted 29.11.19; published 31.01.20.

Please cite as:
Zheng ZW, Yang QL, Liu ZQ, Qiu JL, Gu J, Hao YT, Song C, Jia ZW, Hao C
Associations Between Affective States and Sexual and Health Status Among Men Who Have Sex With Men in China: Exploratory Study Using Social Media Data
J Med Internet Res 2020;22(1):e13201
URL: http://www.jmir.org/2020/1/e13201/
doi:10.2196/13201
PMID:32012054

Original Paper

Health Effects Associated With Electronic Cigarette Use: Automated Mining of Online Forums

My Hua[1*], BA, BSc; Shouq Sadah[1*], PhD; Vagelis Hristidis[1], PhD; Prue Talbot[1], BSc, MSc, PhD

University of California, Riverside, Riverside, CA, United States

[*]these authors contributed equally

Corresponding Author:
Prue Talbot, BSc, MSc, PhD
University of California, Riverside
2320 Spieth Hall
900 University Ave
Riverside, CA, 92521
United States
Phone: 1 9518273768
Email: talbot@ucr.edu

Abstract

Background: Our previous infodemiological study was performed by manually mining health-effect data associated with electronic cigarettes (ECs) from online forums. Manual mining is time consuming and limits the number of posts that can be retrieved.

Objective: Our goal in this study was to automatically extract and analyze a large number (>41,000) of online forum posts related to the health effects associated with EC use between 2008 and 2015.

Methods: Data were annotated with medical concepts from the Unified Medical Language System using a modified version of the MetaMap tool. Of over 1.4 million posts, 41,216 were used to analyze symptoms (undiagnosed conditions) and disorders (physician-diagnosed terminology) associated with EC use. For each post, sentiment (positive, negative, and neutral) was also assigned.

Results: Symptom and disorder data were categorized into 12 organ systems or anatomical regions. Most posts on symptoms and disorders contained negative sentiment, and affected systems were similar across all years. Health effects were reported most often in the neurological, mouth and throat, and respiratory systems. The most frequently reported symptoms and disorders were headache (n=939), coughing (n=852), malaise (n=468), asthma (n=916), dehydration (n=803), and pharyngitis (n=565). In addition, users often reported linked symptoms (eg, coughing and headache).

Conclusions: Online forums are a valuable repository of data that can be used to identify positive and negative health effects associated with EC use. By automating extraction of online information, we obtained more data than in our prior study, identified new symptoms and disorders associated with EC use, determined which systems are most frequently adversely affected, identified specific symptoms and disorders most commonly reported, and tracked health effects over 7 years.

(J Med Internet Res 2020;22(1):e15684) doi:10.2196/15684

KEYWORDS

electronic cigarettes; vaping epidemic; vaping-associated pulmonary illness; e-cigarettes; electronic nicotine delivery devices; health effects; nicotine; symptoms; disorders; pulmonary disease; pneumonia; headaches; content analysis; text classification; e-cigarette, or vaping, product use associated lung injury

Introduction

Background

At the time of their introduction 10 years ago, there was little information on the health effects associated with electronic cigarettes (ECs); nevertheless, they were often considered safer than conventional cigarettes because they do not burn tobacco and therefore produce aerosols with fewer chemicals. Since their introduction, a wide range of studies concerning the health effects associated with ECs have been conducted using various approaches that include online informatics and survey studies [1-6], short-term physiological assessments of EC use on human health [7,8], and *in vitro* and *in vivo* cytotoxicity studies [9-15].

Although these studies are limited mainly to acute exposures, they often suggest that EC use is not harm free [16]. Summaries of the health effect data and case report information on ECs can be found in 2 recent reviews [17,18].

Infodemiological approaches, which mine data from the internet and social media, have yielded new information such as EC topography and the effects of EC use on human health [1,19-22]. For example, in a previous study, we mined internet data on EC puffing topography and showed that puff duration is about twice as long for EC users than conventional smokers [22]. In addition, topography is highly variable among EC users, who generally intake much larger volumes of aerosol than cigarette smokers [23]. In our prior infodemiological study, we mined information manually from major EC online health forums and identified numerous negative and some positive health effects that users attributed to ECs [1]. This was a useful approach; however, manual mining methods are labor intensive, limit the number of posts that can be reasonably extracted and analyzed, and are not practical for examining large amounts of data over time.

Objectives

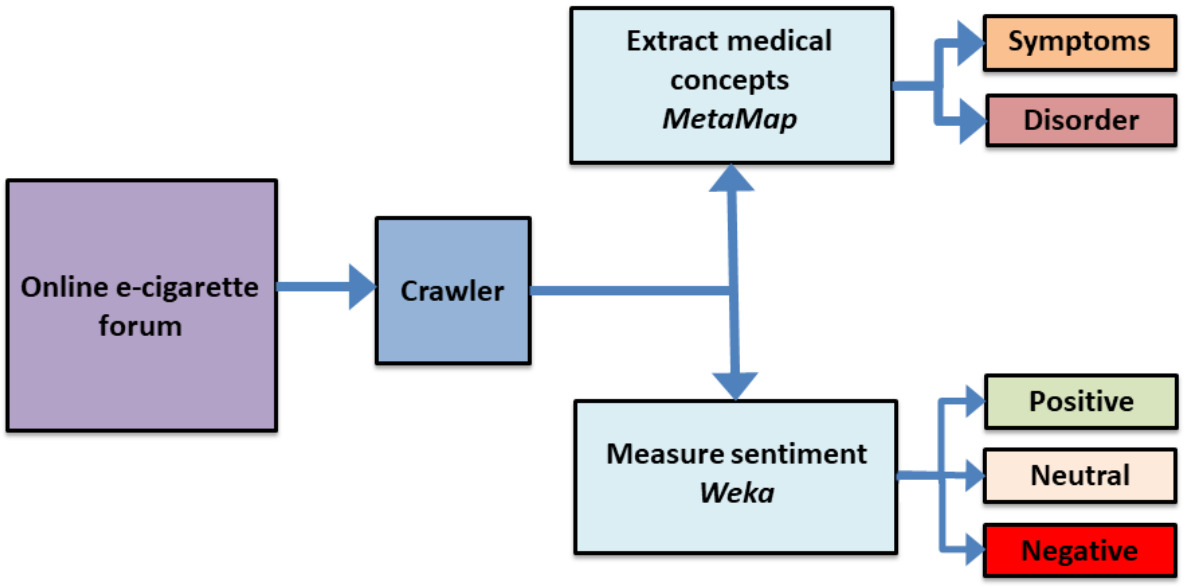
The objective of this study was to use automated computer methods to mine an online forum and extract a large set of posts dealing with the effects of EC use on human health. These data were analyzed to identify the symptoms (undiagnosed conditions) and disorders (physician-diagnosed terminology) associated with EC use. Data were analyzed over a 7-year period, and the sentiment in each post (positive, negative, and neutral) was determined.

Methods

Datasets

We collected data posted between January 2008 and July 2015 on a large EC discussion forum. Data from 2008 and 2015 were each collected for approximately 6 months. We analyzed the layout of the website and built a crawler in Java using the Java library jsoup [24], which is designed to extract and parse information from HTML pages. The posts were collected from 7 subforums. The total number of discussion threads was 2330 and the total number of posts was 1,450,896. As the primary goal of this study was to evaluate the health effects produced by ECs, we focused on those posts that belonged to the 7 health subforums, which contained 41,216 posts. We emphasize that all collected data are publicly available, including discussion threads and users' information. Figure 1 shows the overall pipeline used for our analysis.

Figure 1. Online forum data pipeline showing processing, post sorting, and classification workflow.



Medical Concepts

We used a modified version of the MetaMap tool [25] to annotate each post with a set of medical concepts from the Unified Medical Language System (UMLS). UMLS is a repository of a large number of biomedical controlled vocabularies [26]. In UMLS, there are 15 high-level semantic groups, which were created to help reduce complexity by grouping the semantic types [27]. In this work, we analyzed 2 semantic types, *sign or symptoms* and *disorder or syndrome*, which belong to the *Symptoms* and *Disorders* semantic groups. Each concept in UMLS can be assigned to multiple semantic types, but only to 1 semantic group [27]. As MetaMap was built to annotate the natural language text in biomedical academic publications, it is not very effective out-of-the-box on social media posts, as it successfully maps the medical terms most of the time, and not the descriptive or nonmedical terms [28]. To improve the tool's mapping efficiency, we manually examined and removed misclassified UMLS concepts generated by MetaMap by performing the following steps:

1. For the 2 semantic types (symptoms and disorders), we ordered the concepts by their frequencies.
2. We analyzed the different terms mapped to each concept.
3. We removed the misclassified concepts from our results. Examples of misclassified concepts include:

a. mod, which refers to vape mods, was mapped to Type 2 diabetes mellitus (C0011860)
b. ect, which is a type of vape mod, was mapped to Benign Rolandic epilepsy (C2363129)
c. pic was mapped to Punctate inner choroidopathy (C0730321)

For each semantic type, we reported the most frequent disorders and symptoms overall and by year.

Sentiment

To measure the positive and negative health effects produced by EC use, we used a supervised learning classifier (Random Forest) on a set of manually labeled posts to predict the sentiment for unseen posts. We randomly selected 1080 posts, which were labeled independently by 3 of the authors as the following:

- Negative: if a post clearly contained a health effect or unpleasant experience or complaint that co-occurred with the use of EC.
- Positive: if a post clearly mentioned a health improvement or a recovery from previous health effects when switching from smoking analogs to EC.
- Neutral: if a post did not express any sentiment.

Our interpretation of positive and negative is different from typical sentiment classifications, and mainly focuses on health-related effects. We first asked the labelers to categorize 400 posts, and then we measured the intercoder reliability between the labelers. Using *ReCal* [29], an online tool to calculate the reliability for the masses, the agreement was 80.53% using the *Average Pairwise Percent Agreement* measure. Owing to the high agreement, the rest of the posts were split evenly among the labelers to categorize. Table 1 shows the class distribution of our sample data with examples for each class; 44.7% (179/400) of posts were labeled as negative, 38.5% (154/400) as neutral, and 16.7% (67/400) as positive.

Table 1. Sample data summary.

Class	Posts, n	Example
Positive	180	“I’ve only been vaping for 2 1/2 weeks, but I’ve already noticed a big difference in my lungs (after 20+ years of smoking). For example, I had a chest cold when I started, and in the past, once a cold moved into my chest it took a couple of months to get rid of it. ...E-cigs are pretty darn amazing, IMHO.” [sic]
Neutral	416	“I dont [sic] think there are any tests since flavoring were not meant to be inhaled [sic]. I think we are taking our chances untill [sic] some evidence comes out...”
Negative	484	“Hi Everyone, I have been using e-cigarrette [sic] for the past 2 months and very disappointed [sic] that I have to stop, reason being my teeth, gums are sensitive and my tooth cracked yesterday, I have to have a crown fitted.8-o [sic]. I think that the nicotine is seriously not good for the mouth. My husband and work collegue [sic] have also reported sore gums, little sores in the mouth…”

Using Weka machine-learning toolkit v. 3.8.1 [30], we first filtered our sample data after many experiments using StringToWordVector class filter, which filters strings into N-grams using WordTokenizer class, with the following settings: (1) convert all words to lower case, (2) remove stop words, (3) stem words using Weka built-in stemmer, (4) keep only terms that appear at least twice, and (5) retain unigram, bigram, and trigram. We then split the sample data as follows: (1) 962 posts for the training test and (2) 118 posts for the test set. We then trained our data using the Random Forest classifier; however, the classifier’s initial accuracy was not satisfactory.

To improve the classifier’s accuracy, we needed to address a well-known issue in our sample data, which is the imbalanced class distribution [31]. The Positive class, as seen in Table 1, only covers 16.7% (67/400) of the data, whereas the Neutral class covers 38.50% (154/400) and the Negative class covers 44.7% (179/400). Thus, we oversampled the Positive class by duplicating the posts which were labeled Positive in the training set only. Table 2 shows the new class distribution for the training set, namely Training (extended). Another approach we used to improve the accuracy is annotating all the posts in the sample data with the ancestors of the medical concepts mentioned in the posts. For example, if *pneumonia* is mentioned in a post, then we append with *Disorder of lung*.

After using the new training set, the classifier’s accuracy increased from 66.95% to 75.42%. Table 3 reports for each class 3 different measures, including precision, recall, and F-measure. As seen in the table, the classifier is most accurate on the Negative class (F-measure=0.79), followed by Positive and Neutral classes.

Table 2. Training data summary (N=400).

Class	Training, n (%)	Training (extended), n (%)
Positive	67 (16.75)	112 (28.0)
Neutral	154 (38.50)	136 (34.0)
Negative	179 (44.75)	152 (38.0)

Table 3. Test data classification accuracy (N=118).

Class	Precision	Recall	F-measure	Posts, n
Positive	0.73	0.72	0.74	21
Neutral	0.67	0.77	0.71	39
Negative	0.84	0.74	0.79	58
Average	0.76	0.75	0.76	118

Data Categorization and Analysis

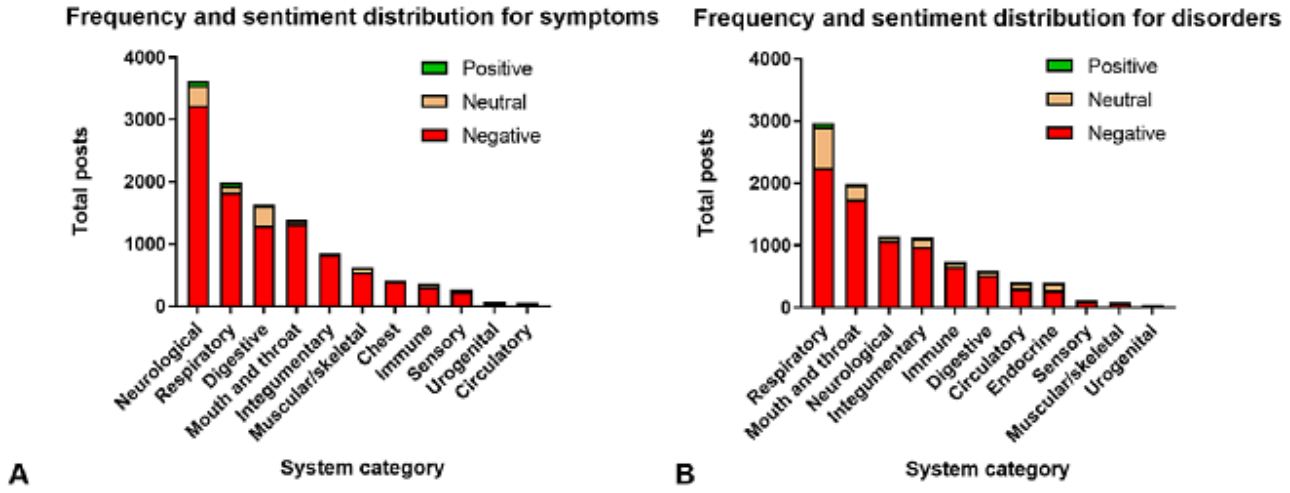
All health-related effects (symptoms and disorders) data reported by EC users in posts were collected iteratively and sorted into Microsoft Excel spreadsheets. The symptoms and disorders were further grouped according to the organ system and anatomical region, which we defined as *systems* previously [1]. When a symptom could have been associated with more than 1 system, the health effect was assigned to the system for which it had the strongest fit (eg, improved sense of taste was assigned to sensory but could have been mouth/throat). Frequency distributions for the overall grouped data in each system for symptoms and disorders were plotted using GraphPad Prism (GraphPad, San Diego). In addition, the sentiment for each post was grouped according to their positive, neutral, and negative sentiment as described in the Methods section.

Results

Overall Frequency of Reported Symptoms and Disorders Classified by System or Anatomical Region

The 41,216 posts we collected spanned the years from 2008 to 2015 (2008 and 2015 were half years). We analyzed the frequency of reports for various symptoms and disorders by consolidating the reported health effects into structural or physiological systems (eg, sore throat was classified into mouth and throat; Figure 2). The 5 systems that had the most reports for symptoms were neurological (n=3623), respiratory (n=1995), digestive (n=1637), mouth and throat (n=1390), and integumentary (n=853; Figure 2). The top 5 systems for disorders were respiratory (n=2972), mouth and throat (n=1986), neurological (n=1143), integumentary (n=1123), and immune (n=739; Figure 2). For both symptoms and disorders, a majority of the posts were associated with negative sentiment across all systems (Figure 2).

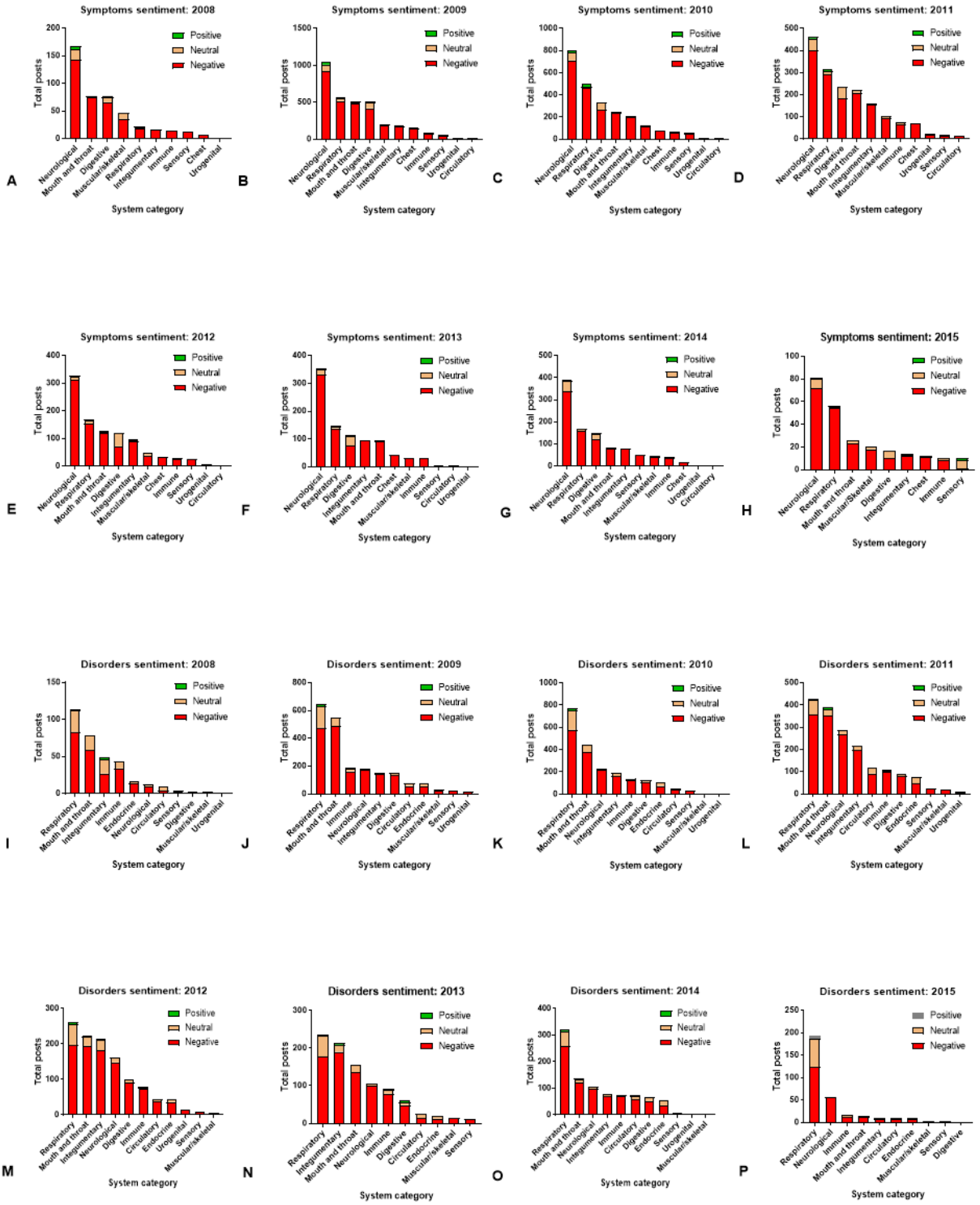
Figure 2. Frequency distribution of reported symptom (A) and disorder (B) posts grouped into their systems or anatomical regions. The frequency of positive, neutral, and negative posts is shown for symptoms (A) and for disorders (B).



Symptom and Disorder Frequency and Sentiment Distribution Over Time

After examining overall frequency distribution for all posts, we grouped the posts according to their years for analysis in their symptom or disorder categories. Across all years for both symptoms and disorders, we found the frequency distribution of reports per year. In addition, the posts for symptoms and disorders were categorized according to sentiment (positive, negative, and neutral), and their frequency per year was summarized in stacked bar graphs for each year (Figure 3).

Figure 3. The frequency distribution of positive, neutral, and negative sentiment was assigned for reported symptoms (A-H) and disorders (I-P).



For the symptoms, the posts with the most reports were consistently found in the neurological, respiratory, digestive, integumentary, and mouth and throat systems. For all years except 2008, the neurological and respiratory systems were the top 2 systems. The digestive, integumentary, and mouth and throat alternated in some years, but were generally in the top 5 systems with the most posts in each of the years.

Similarly, the posts containing disorders associated with EC use had similar results for their top 5 system categories across the 7 years of reporting. The 2 top systems reported between 2008 and 2012 were the respiratory and mouth/throat. Alternating in the top 5 disorders were the integumentary, neurological, and immune systems.

Negative sentiment was associated with most symptoms and disorders in each system or anatomical region (Figure 3), with some increases reported in positive health effects in 2015 for the disorders (Figure 3). It should also be noted that we only

have partial reporting for 2015 because data collection was terminated by the EC forum.

Specific Symptoms and Disorders in Systems With the Most Reports

Heat maps were made by plotting the frequency with which individual symptoms/disorders occurred for all 41,216 posts (Figures 4 and 5; Multimedia Appendices 1 and 2). Symptoms with fewer than 2 posts are listed in Multimedia Appendix 3. The total number of negative and positive posts for each symptom or disorder is shown on a log scale ranging from high (red) to low (blue). White represents a zero-post frequency. Numerous negative symptoms were reported for each system. Typically, about 16.52% (6807/41,216) of the symptoms were reported frequently (red), while the majority often occurred in fewer than 100 posts (blue to purple). In the neurological system, the most common symptoms included: headaches (n=939), fatigue/tired/malaise (n=468), nausea (n=290), dizziness (n=183), and lightheadedness (n=113; Figure 4). In the respiratory system, the negative effects included: coughing (n=852), wheezing (n=298), dyspnea (n=235), and excessively deep breathing (n=112). The most reported symptoms in the digestive system were: heartburn (n=327), cramping (n=303), flatus (n=176), and constipation (n=113). In the mouth/throat and integumentary systems, common symptoms were: pain in throat (n=643), harsh voice quality (n=175), pharyngeal dryness (n=147), itching skin (n=565), and dry skin (n=121; Figure 4). Other commonly reported symptoms involved aching and chest pains as well as immune symptoms related to the cold and flu.

Although positive symptoms were not frequently reported in this online forum, those reports that were posted most often dealt with improvements in the neurological (n=77), respiratory (n=60), digestive (n=19), and mouth and throat (n=18) systems (Figure 4). In the neurological system, these include improvement in tiredness (n=12) and insomnia (n=8). For respiratory system, these symptoms included improvements in wheezing (n=17), dyspnea (n=14), and coughing (n=8). In the digestive and mouth and throat systems, improvements were found in cramp (n=5) and halitosis (n=5). Other systems and anatomical regions had fewer than 10 total positive reports.

For each system/anatomical region, there were 1 to 3 top disorders. In the respiratory system, the most common disorders were asthma (n=916), chronic obstructive pulmonary disorder (COPD; n=471), pneumonia (n=367), and bronchitis (n=232; Figure 5). In mouth and throat, the most common disorders were pharyngitis (n=565), aptyalism (n=377), and ulcer of mouth (n=207). The most reported disorders in the neurological system were dehydration (n=403) and migraine (n=103). Most disorders were reported in the respiratory, mouth and throat, neurological, integumentary, and immune systems (Figure 5), whereas the remaining systems had fewer reported disorders (Multimedia Appendix 2).

Figure 4. Heat map of specific symptoms reported in the neurological, respiratory, digestive, mouth and throat, and integumentary systems. The total number of posts for each symptom is shown on a log scale ranging from high (red) to low (blue).

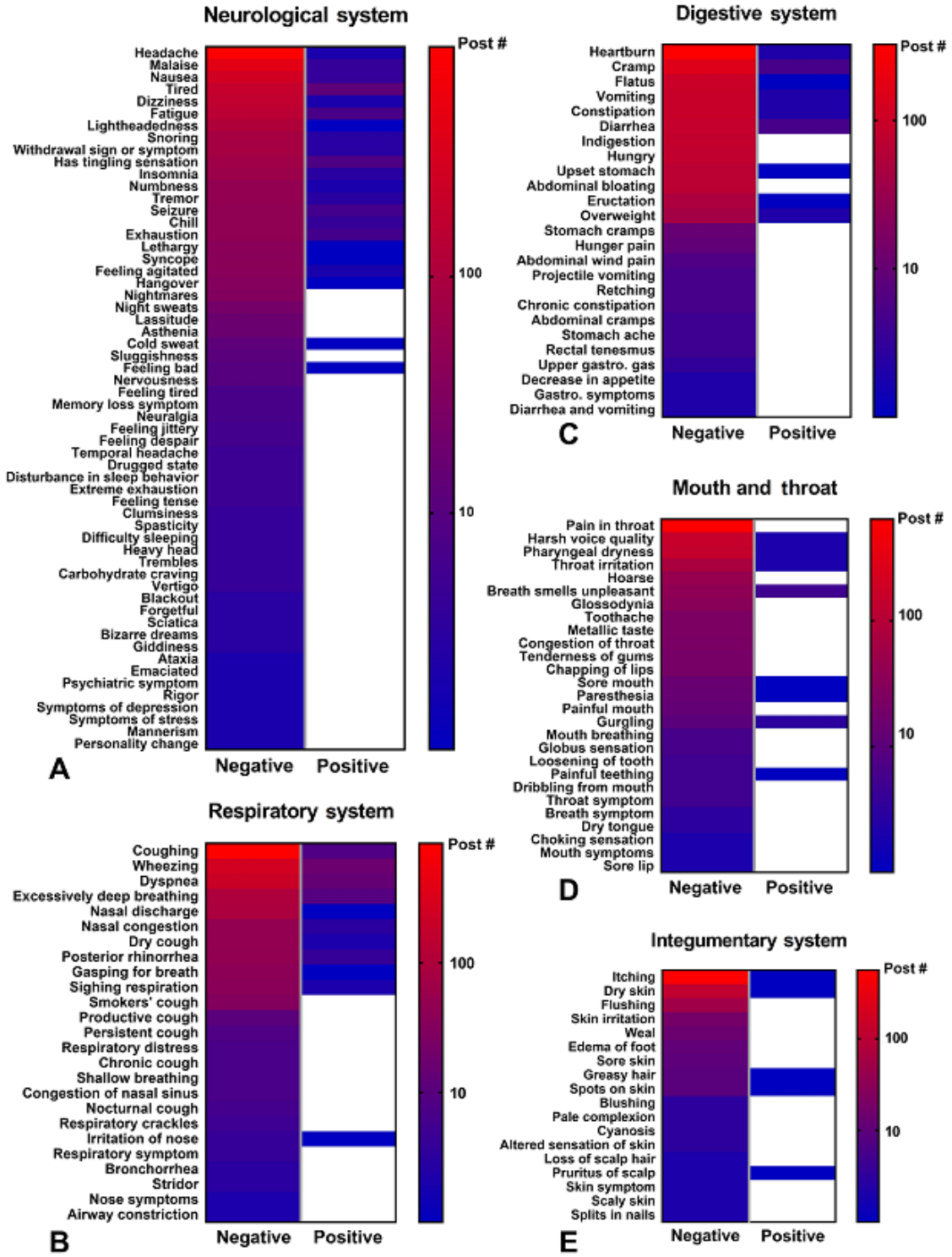
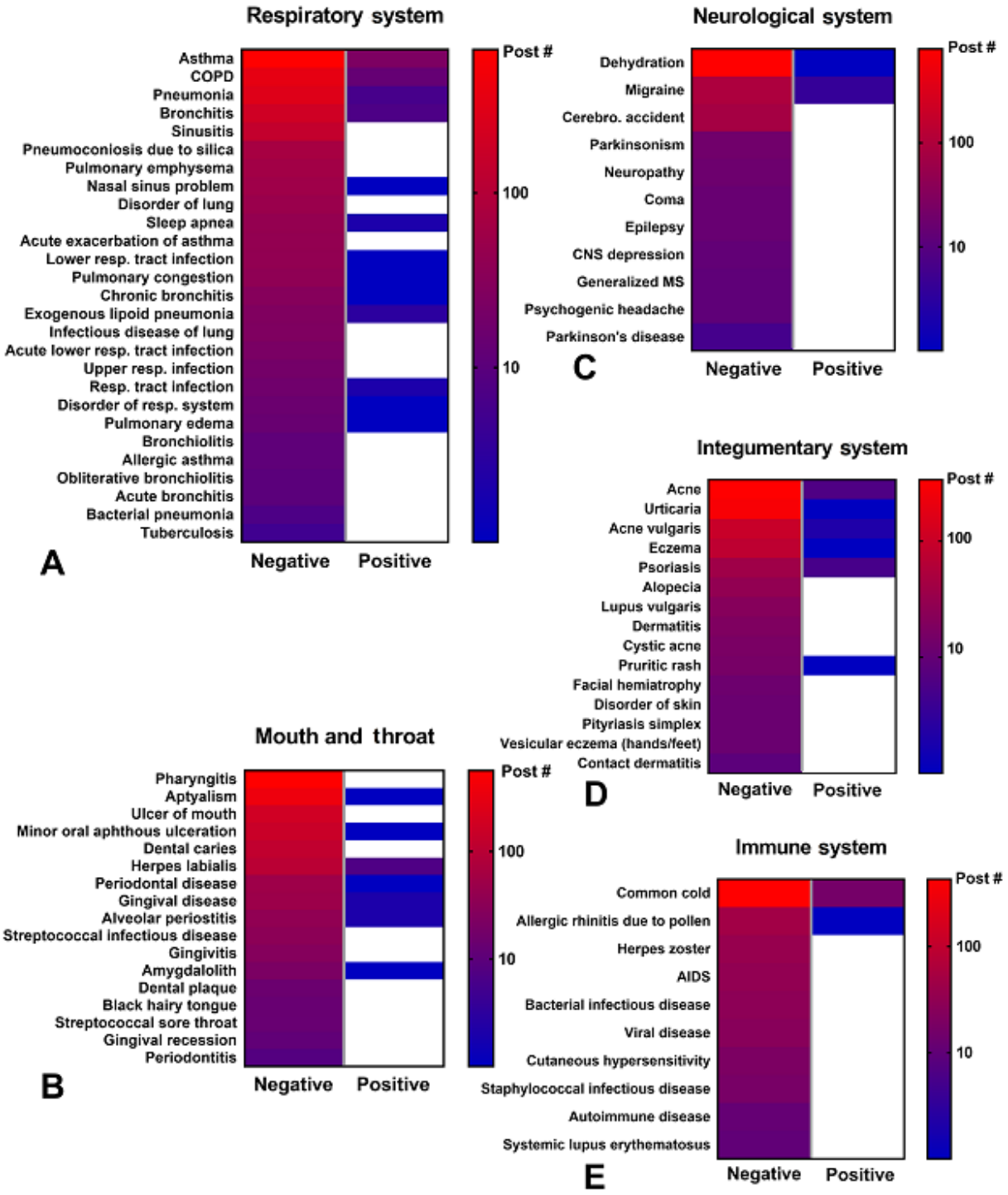


Figure 5. Heat map of specific disorders reported in the respiratory, mouth and throat, neurological, integumentary, and immune systems. The total number of posts for each disorder is shown on a log scale ranging from high (red) to low (blue).



To compare the frequency with which different symptoms/disorders appeared across different systems, frequency distribution graphs were created (Figures 6 and 7). Graphs show only those symptoms/disorders with over 100 posts (Figures 6 and 7). These data were sorted by negative sentiment as negative effects were most commonly reported and were of most interest. In total, 25 symptoms and 22 disorders had over 100 posts. The 5 top symptoms in the 41,216 posts were: headache, coughing, pain in throat, itching, and malaise (Figure 6). The top 5 disorders in the dataset were dehydration, asthma, pharyngitis, common cold, and aptyalism (Figure 7). These symptoms and disorders are the most commonly reported conditions in our dataset.

Figure 6. Frequency distribution of specific symptoms with over 100 posts and frequency distribution of their systems or anatomical regions (inset). Digest.: digestive; Integ.: integumentary; Mo./Th.: mouth and throat; Musc./Skel.: muscular/skeletal; Neuro.: neurological; Resp.: respiratory.

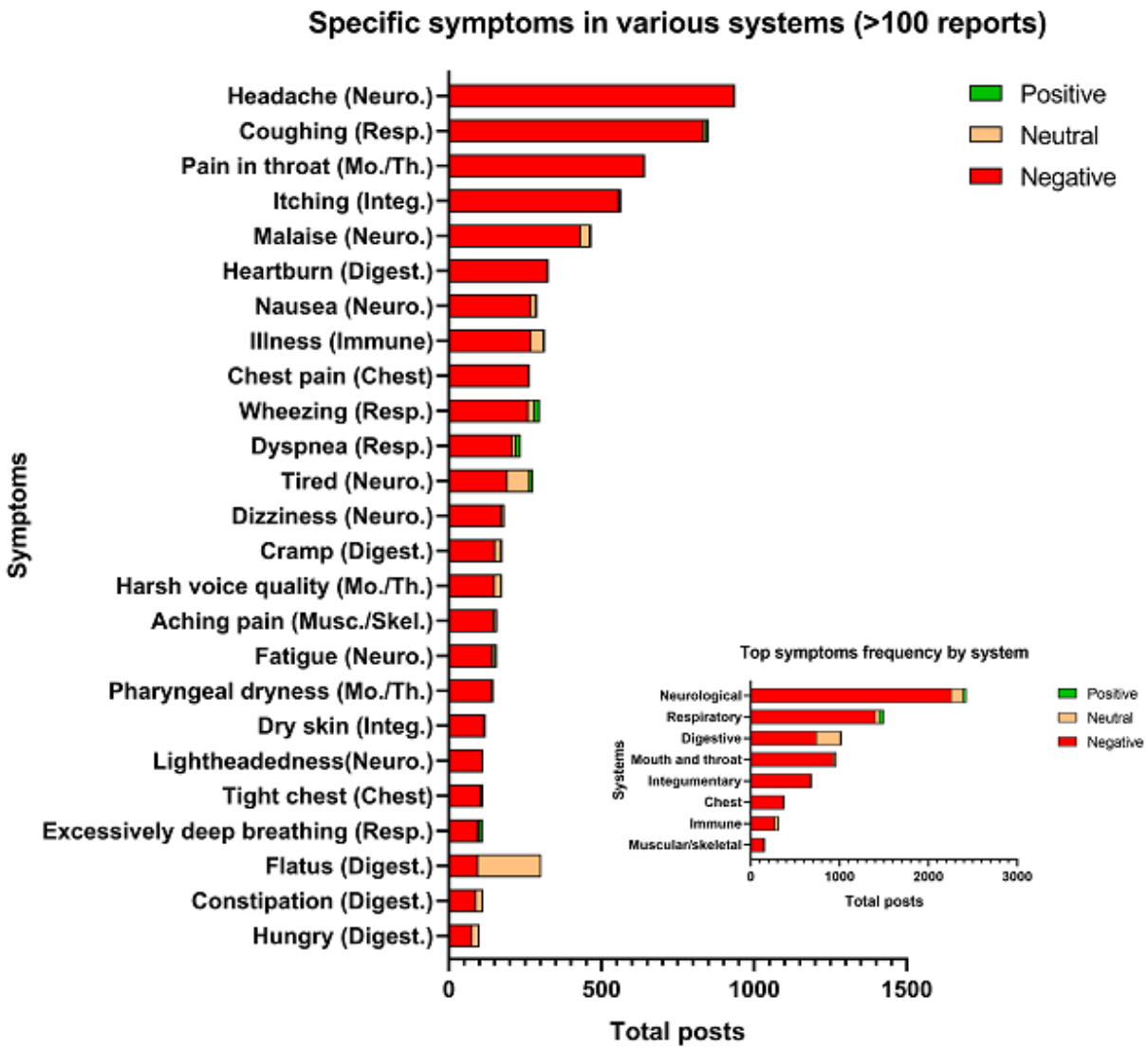
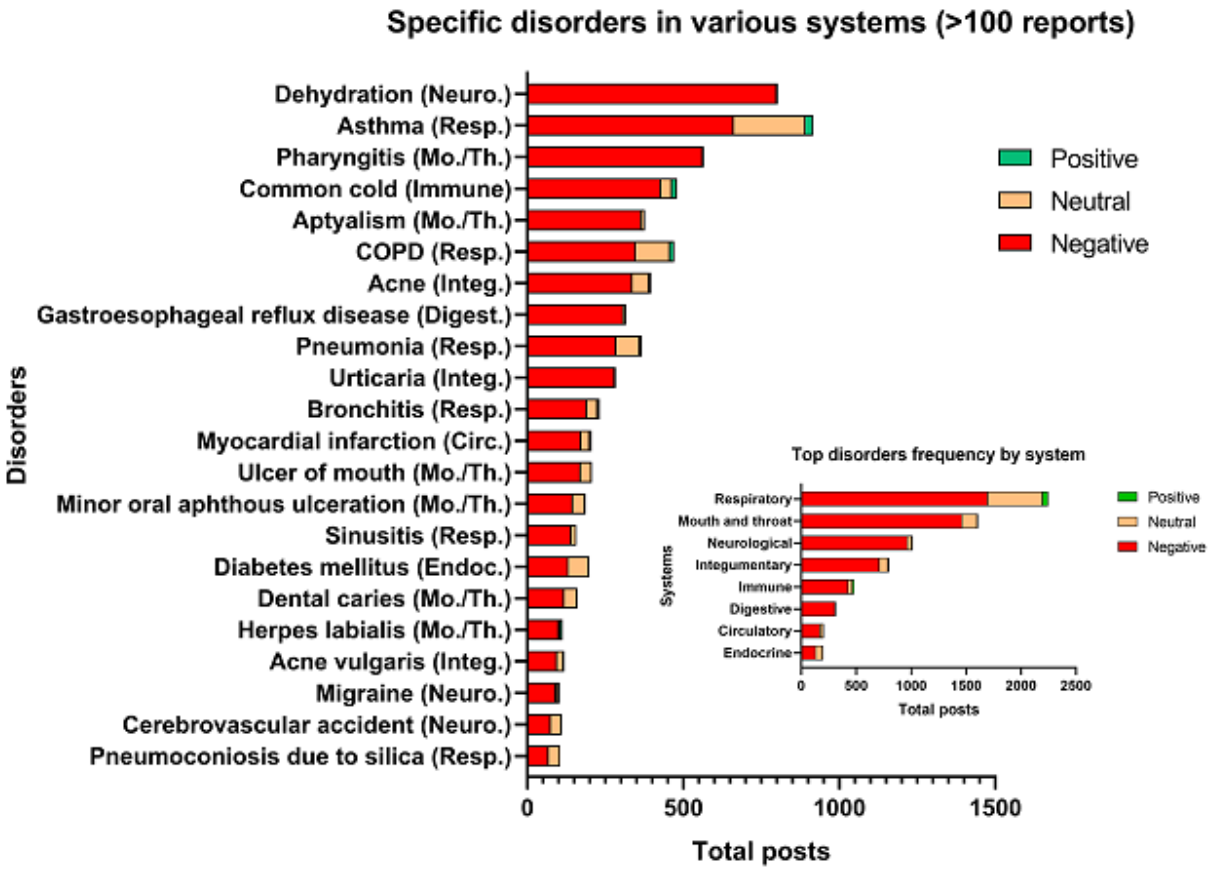


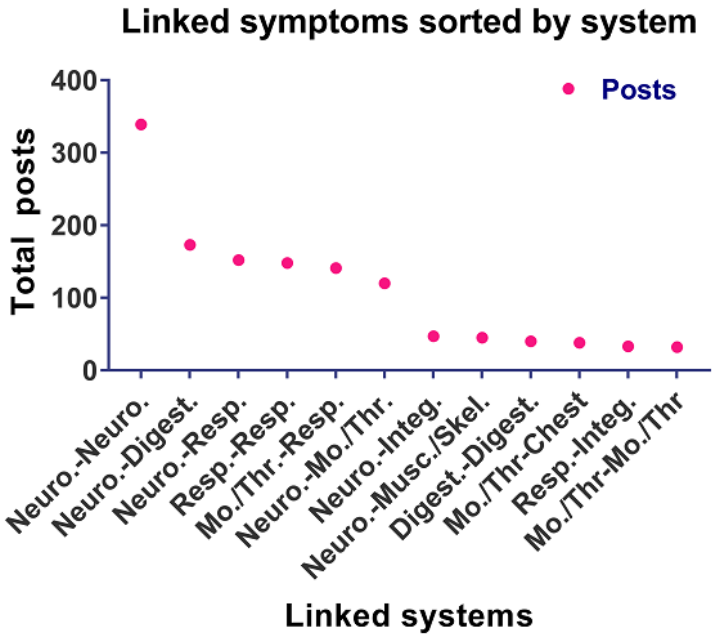
Figure 7. Frequency distribution of specific disorders with over 100 posts and frequency distribution of their systems or anatomical regions (inset). Neuro.: neurological; Resp.: respiratory; Mo./Th.: mouth and throat; Integ.: integumentary; Digest.: digestive; Circ.: circulatory; Endoc.: endocrine.



Identification of Frequently Reported Paired Symptoms

A total of 46 paired symptoms were frequently reported (Figure 8; Multimedia Appendix 4). Those with over 30 reports included a combination of neurological-neurological symptoms (eg, nausea and headache), respiratory-respiratory symptoms (eg, wheezing and coughing), and/or neurological-respiratory-mouth and throat symptoms (eg, pain in throat and headache; coughing and headache). The results in the top symptom pairings reflect the abundance of symptoms reported in their respective categories. As for those pairings occurring in less than 30 posts, various combinations of neurological, respiratory, integumentary, and digestive related symptoms.

Figure 8. Graph showing frequency with which symptoms in various systems were linked. Digest.: digestive; Integ.: integumentary; Mo./Th.: mouth and throat; Musc./Skel.: muscular/skeletal; Neuro.: neurological; Resp.: respiratory.



Discussion

Principal Findings

The internet is a dynamic resource containing information that can be mined to learn about the health effects associated with EC use. In our previous study, we manually mined 632 posts from 3 online EC forums to identify both positive and negative health effects reported by EC users [1]. Manual mining of such information is time consuming, labor intense, and limited by the number of posts that can be realistically evaluated. To take additional advantage of the internet as a repository of EC-related health information, we developed an automated method that was used to extract over 1 million posts from an EC forum. These posts were then filtered to yield over 41,000 health-related posts for detailed analysis. By automating the extraction process, we collected 100 times more posts for analysis and tracked responses over 7 years, an option that would be too time consuming to perform manually. The data showed a variety of positive and negative symptoms/disorders, demonstrating that the internet is a valuable resource for acquiring new data related to EC usage and their associated health effects.

The results from this study are in overall good agreement with our prior publication [1]. In both studies, the neurological and respiratory systems were most often reported to have adverse effects associated with EC use. Both studies reported similar positive and negative health effects in various systems, and within all systems, a small number of self-reported symptoms and disorders occurred at high frequency (Figures 6 and 7). However, with the power of automated mining, we were also able to (1) identify numerous symptoms and disorders, many of which were not previously reported, along with their frequencies; (2) report new data in the disorders category for each system; (3) identify those symptoms/disorders that users have reported most frequently which will be of interest to health care providers treating patients using EC products; (4) show that health effects were similar over a 7-year period; (5) evaluate positive and negative sentiments for symptoms and disorders; (6) evaluate symptoms that are linked to each other; and (7) identify top symptoms (eg, wheezing) and disorders (eg, asthma, COPD, and pharyngitis) that are associated with inflammation. In addition, some symptoms (eg, headache and nausea) and disorders (eg, pneumonia) that occurred with the highest frequency in the neurological and respiratory system have also appeared in a number of case reports in the EC literature [17,32-35]. Also, in agreement with our prior study, some reported health outcomes attributed to EC use were positive [1]. These included reduction in symptoms such as coughing and wheezing and disorders such as asthma, COPD, and common cold. It is very likely that there were real health benefits for some individuals, especially for those switching from conventional to EC use, and this is supported by other publications [36-38].

There are numerous reports on the health effects of EC, many of which are in agreement with our data. In the neurological system, the most commonly reported adverse symptoms we observed included headache, fatigue, nausea, dizziness, and seizures, which have also been reported in human studies [39-42]. Headaches have been reported to the Food and Drug Administration (FDA) by EC users [43], they were a common side effect in various surveys and online studies [4,44-46] and were reported in human studies in which participants used different EC devices and refill fluids with varying nicotine concentrations [47-49]. In 1 case report, an adult male experienced severe headaches/migraines and seizures for 1 week before being diagnosed with reversible cerebral vasoconstriction syndrome related to EC use [32], and in a second case, an adolescent female developed persistent daily headaches after a single EC use [50]. Nausea and dehydration were commonly reported symptoms and disorders of the neurological system.

These symptoms have been associated with headache and fatigue/tiredness [51-54]. In our study, headache and nausea were frequently reported together, demonstrating that symptoms associated with EC use may be linked. The frequency distribution showing symptoms and disorders with over 100 posts also revealed that digestive symptoms such as heartburn and circulatory disorders such as myocardial infarction were highly reported. Digestive symptoms related to EC use have not been previously focused on and may be important to pay attention to, and disorders such as myocardial infarction associated with EC have recently received more attention from epidemiology studies [55].

For the respiratory system, the most frequently reported symptoms included coughing, wheezing, and dyspnea, and the top paired respiratory symptoms were coughing-wheezing. In the national Population Assessment of Tobacco Health (PATH) and in some human surveys, EC use was associated with increased wheezing (an important potential risk factor for respiratory disease) [56-59]. In our study, the top disorders were asthma, COPD, pneumonia, bronchitis, and sinusitis, which have a common theme of inflammation. Human studies and surveys have shown that adolescents and adults associated chronic bronchitis symptoms (eg, cough, phlegm, or dyspnea) with EC use [57,58]. Epidemiological studies have linked EC use to both COPD and asthma [56], and the PATH study showed that dual use of EC and conventional cigarettes aggravated this risk [56]. Frequently reported respiratory disorders in our study such as pneumonia and bronchitis have appeared in several EC case reports, most of which deal with lung inflammation and pneumonia-linked incidents [34,35,60,61]. Some of the patients in these case reports had no preexisting health conditions but presented with coughing, wheezing, and dyspnea. They typically recovered from their respiratory disorders after discontinuing EC use.

The circulatory, mouth/throat, chest, integumentary, and immunological systems were also affected by EC use in our study. Symptoms such as pain in throat, dry skin, pounding heart, and chest pain have been reported in survey/human EC studies [62]. Myocardial infarction, which was a top disorder for the circulatory system in our data, has been described in a case report after the patient used an EC with high nicotine [63]. A link between EC use and myocardial infarction was also found in a recent national survey adjusted for conventional smoking and other risk factors [55]. EC users are potentially susceptible to periodontal disease and increased plaque formation, which could lead to dental caries (also reported in our paper) [64]. Other reported disorders in our study, such as common cold and diabetes mellitus, are immunologically based, and multiple studies have shown EC aerosol exposure can induce inflammatory response [65]. Mice exposed to EC aerosol have impaired pulmonary viral and bacterial clearance ability that can lead to increased bacterial resistance, implying that EC users are more susceptible to cold and flu, a common complaint in our online studies. Experimental studies have further demonstrated that EC aerosol exposure can result in oxidative stress [13,66,67], suggesting chronic use could trigger inflammation, leading to progressive inflammatory disorders in the respiratory system.

Electronic Cigarette Aerosol Chemicals That May Produce Adverse Health Effects

EC refill fluids and aerosols are complex mixtures that contain flavor chemicals, solvents, nicotine, and metals that could contribute to adverse health effects (Table 4). Although most flavor chemicals in EC are generally regarded as safe (GRAS) for ingestion, their inhalation safety has usually not been established [68]. Some EC products contain high concentrations of flavor chemicals that exceed the National Institute of Occupational Safety and Health limits [69-75] and the concentrations normally used in consumer products [71,72]. Many EC flavor chemicals are classified as irritants [72] and are cytotoxic when tested *in vitro* at concentrations below those in EC products [71,72]. Cinnamaldehyde, which is used in refill fluids including those that do not have cinnamon in their name [10], is highly cytotoxic in the 3-(4,5-dimethylthiazol-2-yl) -2,5-diphenyltetrazolium bromide [10,69] and impaired respiratory response in immunological assays [76-78] and *in vivo* assays. In cell, animal, and human studies, EC flavor chemicals (eg, citrus/fruit and chocolate) caused an increase in reactive oxygen species leading to tissue and DNA damage and inflammation [67], which could in turn lead to mutations and disease progression. Some flavor chemicals, such as alcohols and phenols, can dilate blood vessels and cause headache, nausea, and fatigue. Prolonged inhalation of flavor chemicals, such as benzaldehyde, ethyl butanoate, diacetyl and its derivatives (2,3-pentadione, acetoin), triacetin, and limonene can elicit headaches, dizziness, and/or respiratory symptoms. Diacetyl (2,3-butanedione), a diketone associated with respiratory symptoms (wheezing and shortness of breath) and bronchiolitis obliterans, is in some EC refill fluids and can form as a reaction product during aerosolization [70,74]. Additional reaction products that form in EC aerosols (eg, aldehydes, acetals, and oxides) can be harmful to humans and elicit various symptoms, including pain.

Table 4. Examples of chemical components in electronic cigarettes that may cause major symptoms/disorders with reference citations of studies.

Symptom/disorder	System	Flavor chemicals (study)	Metals (study)	PG[a]/VG[b]/byproducts (study)	Nicotine (study)
Headache	Neurological	[79,80]	[81-83]	[51,52,84,85]	[86-88]
Fatigue/malaise	Neurological	[89]	[90]	—[c]	—
Dizziness	Neurological	[91]	[92,93]	—	[94]
Nausea	Neurological	[79,80]	[92]	[85,95]	[87,96]
Dehydration	Neurological	—	—	[52,84]	[97]
Coughing	Respiratory	[91,98-100]	[62]	[101,102]	—
Wheezing	Respiratory	[57,99]	[103]	[104]	—
Dyspnea	Respiratory	[99]	[105,106]	[107]	[108]
Asthma	Respiratory	[99,109]	[92,103,104]	[110]	—
COPD[d]	Respiratory	[111]	[92]	[110]	—
Pneumonia	Respiratory	—	[103]	—	—
Bronchitis	Respiratory	[111,112]	[113]	—	—
Sinusitis	Respiratory	[79,114]	[115]	[114]	—
Pain in throat	Mouth and throat	—	—	[101]	[87]
Dental caries	Mouth and throat	[116]	[117]	—	[118]
Itching/urticaria	Integumentary	—	[119-121]	[122]	[123]
Dry skin	Integumentary	—	—	[124]	—
Acne	Integumentary	—	[125]	—	[126,127]
Heartburn	Digestive	—	—	—	[128]
Cramp	Digestive	—	[129]	—	—

[a]PG: propylene glycol.

[b]VG: vegetable glycerin.

[c]Lack of evidence in referenced literature.

[d]COPD: chronic obstructive pulmonary disorder.

Elements/metals (eg, aluminum, copper, cadmium, chromium, iron, nickel, silicon, lead, cobalt, and zinc) have been identified in EC aerosols [130,131]. In studies not involving ECs, these elements/metals have been linked to neurological (headache, nausea, and dizziness) and respiratory (eg, coughing, wheezing, shortness of breath, and bronchial/pulmonary irritations impairment) symptoms [132]. A positive correlation has been reported between human EC use and internal concentrations of nickel and chromium [133]. We found *pneumoconiosis due to silica* frequently reported by users, which could be caused by silica particles in EC aerosols [130]. Inhalation of silicon particles can elicit cough, inflammation, and lung fibrosis [134]. Although there is not a consensus on whether element/metal concentrations in EC aerosols are high enough to produce these effects [135], some evidence suggests that they could be a factor [130,136]. A female patient with no history of allergic disease tested positive for nickel allergy after being diagnosed with dermatitis caused by corrosion of the EC device [119]. Some of the top symptoms and disorders in our study also relate to inflammation of the skin (eg, itching and eczema), and this may be attributed to direct exposure and allergic reactions to EC products [119,120].

Propylene glycol and glycerin, 2 solvents in EC aerosols, are generally considered safe for ingestion; however, they are known respiratory tract and integumentary irritants [137]. When heated, propylene glycol can produce toxic aldehyde reaction products, such as acetaldehyde and formaldehyde [138], which can cause cellular and tissue damage in the body [67]. Inhalation of aldehyde fumes can cause dizziness, nausea, and headaches in humans [113], and formaldehyde can cause coughing, wheezing, pneumonia, bronchitis, and neurological and cardiovascular symptoms (eg, headaches, nausea, heart palpitations) [139]. Inhalation of propylene glycol mists can elicit both neurological and respiratory symptoms, such as nausea, wheezing, shortness of breath, and cough [101,140] and can exacerbate asthma [101,140]. Propylene glycol and glycerin produce 15 different aerobic thermal degradation products through hydrogen abstraction, oxidation, and cleavage reactions [138]. Several of these (eg, formaldehyde, formic acid, acetaldehyde, and acrolein) are carcinogens or have genotoxic potential [141,142]. Some of these byproducts are hemiacetals (such as formic acid and formaldehyde), which equilibrate and persist in the aerosols inhaled by the users.

Nicotine, a major component in most EC fluids, has various neurological, respiratory, digestive, mouth/throat, and circulatory

system effects that overlap the symptoms/disorders observed in our study. Most cases of EC nicotine poisonings result from oral ingestion or intravenous injection [17] and are characterized by symptoms such as vomiting, nausea, dizziness, headaches, and more severe effects that can lead to death. The side effects of nicotine inhalation include headache, nausea, mouth/throat pain, cough, and heartburn [143]. Some users in our study may have been weaning themselves off nicotine or using devices with poor nicotine delivery leading to nicotine withdrawal, which could produce symptoms such as dizziness and anxiety [144]. Nicotine can trigger a dose-dependent loss of the endothelial barrier which has been shown to rapidly increase lung inflammation and oxidative stress in mice [15]. Nicotine in EC aerosols can induce glucose deprivation in the brain, which could lead to enhanced ischemic brain injury and or stroke risk. In addition, EC refill fluids may contain free-base nicotine (a form more addictive) [145], which can lead to greater deposition in the mouth and throat and upper respiratory tract [146].

Recently, an e-cigarette, or vaping, product use associated lung injury (EVALI) epidemic has been identified by the Centers for Disease Control and Prevention (CDC) [147]. As of December 2019, at least 2409 cases of lung injury have been reported to the CDC from 50 states, the District of Columbia, and 2 US territories [147]. In addition, 52 deaths associated with vaping were confirmed by 26 states and the District of Columbia [147]. Some of the commonly reported symptoms in presenting patients included chest pain, shortness of breath, cough, nausea, vomiting, diarrhea, fever, chills, fatigue/malaise, and headache [148-150], all of which are reported in this study. In addition to lung-related disease, some case reports included neurological and gastrointestinal symptoms [148,150], which overlap those found in our study.

The sudden uptick in health-related symptoms and conditions related to vaping comes at least 10 years after the products have gained widespread popularity in the United States, including the rise in popularity of JUUL and marijuana (THC) vape products. Our data show that many of the symptoms characterizing the current patients have been reported online for at least 7 years, suggesting that cases similar to those in the current epidemic have existed previously and been unreported or not linked to vaping. Our data further suggest that this epidemic will continue to grow given the many reports of symptoms characteristic of EVALI on the internet. The specific causes of the reported health effects are not yet known, but it is important to continue vigilant reporting of cases, tracking symptoms, and ongoing research on the health effects related to EC use to understand and contain the vaping epidemic.

Limitations

Our data may underestimate positive health effects, which EC users are less likely to post on online forums. The factors causing the symptoms and disorders reported by EC users could be complex and will require further investigations. Demographic data on the study population were not extractable. It is not known if any individuals were dual users or if they had preexisting health conditions that may have affected their response to EC.

Conclusions

This study is the first to use automated methods to analyze posts on an EC website over a span of 7 years and to identify the symptoms and disorders most frequently reported online by EC users. We demonstrate the value of using automated methods to acquire and analyze large datasets thereby increasing the power of infodemiological analyses. In addition, from our dataset, we identified a condensed list of symptoms and disorders and ranked them according to post frequency. These symptoms and disorders reported in our study may be of interest to physicians and health care providers who are treating patients using EC and could potentially be reported more frequently by EC users. Moreover, informative data were collected from a large population of EC vapers irrespective of their EC products and individual topographies and was not limited to a small selection of EC products or human subjects, as is often the case with experimental studies and case reports. Data collected using our automated method contribute to the growing body of knowledge linking EC use to adverse health effects, mainly in the mouth and throat and the neurological, respiratory, digestive, and integumentary systems. Our study identified hundreds of negative effects that were not previously described in case reports and peer-reviewed literature. The results from our study are in good agreement with previous surveys, human studies, and case reports. Although many of the symptoms that were reported with high frequency are not life-threatening (eg, headache, coughing, heartburn, sore throat), they can be disabling and reduce the quality of life. Of particular concern are the respiratory disorders that appeared with high frequency, such as asthma, COPD, pneumonia, and bronchitis, which not only severely impact the quality of life but may also be life threatening. Our data support the idea that EC use is not free of adverse health effects and that it is important to continue tracking the health of EC users. Advances in internet data mining provide a novel method for monitoring the health of EC users over time. Infodemiological data gathered on EC users will be valuable to physicians, regulatory agencies, and the users themselves.

Acknowledgments

The research reported in this publication was partially supported by the National Institute of Drug Addiction, National Institute of Environmental Health Sciences, and the FDA Center for Tobacco Products (CTP) Grant #s R01DA036493 and R01ES029741 (CTP) to PT. MH was supported in part by a Cornelius Hopper Fellowship and Predoctoral Dissertation Fellowship from the Tobacco-Related Research Program of California (Grant # 28DT-0009). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors are grateful to Malcolm Tran, Jacqueline Wong,

and Avni Parekh for their help in organizing and sorting the data. They would like to thank Kevin Huang for his help with the manuscript and references and Dr Duc Nguyen for reading the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Figure 1: Heat map of symptoms less frequently reported in forums along with their associated systems. The total number of posts for each symptom is shown on a log scale.
[PNG File , 109 KB - jmir_v22i1e15684_app1.png]

Multimedia Appendix 2
Supplementary Figure 2: Heat map of disorders less frequently reported in forums along with their associated systems. The total number of posts for each disorder is shown on a log scale.
[PNG File , 86 KB - jmir_v22i1e15684_app2.png]

Multimedia Appendix 3
Supplementary Table 1: Listing of all symptoms not included in heat map.
[DOCX File , 15 KB - jmir_v22i1e15684_app3.docx]

Multimedia Appendix 4
Supplementary Table 2: Listing of all linked symptoms by frequency.
[DOCX File , 14 KB - jmir_v22i1e15684_app4.docx]

References

1. Hua M, Alfi M, Talbot P. Health-related effects reported by electronic cigarette users in online forums. J Med Internet Res 2013 Apr 8;15(4):e59 [FREE Full text] [doi: 10.2196/jmir.2324] [Medline: 23567935]
2. Polosa R, Morjaria JB, Caponnetto P, Campagna D, Russo C, Alamo A, et al. Effectiveness and tolerability of electronic cigarette in real-life: a 24-month prospective observational study. Intern Emerg Med 2014 Aug;9(5):537-546. [doi: 10.1007/s11739-013-0977-z] [Medline: 23873169]
3. Ayers JW, Leas EC, Allem J, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017;12(3):e0170702 [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]
4. Li Q, Zhan Y, Wang L, Leischow SJ, Zeng DD. Analysis of symptoms and their potential associations with e-liquids' components: a social media study. BMC Public Health 2016 Jul 30;16:674 [FREE Full text] [doi: 10.1186/s12889-016-3326-0] [Medline: 27475060]
5. Soule EK, Nasim A, Rosas S. Adverse effects of electronic cigarette use: a concept mapping approach. Nicotine Tob Res 2016 May;18(5):678-685 [FREE Full text] [doi: 10.1093/ntr/ntv246] [Medline: 26563262]
6. Chen Z, Zeng DD. Mining online e-liquid reviews for opinion polarities about e-liquid features. BMC Public Health 2017 Jul 7;17(1):633 [FREE Full text] [doi: 10.1186/s12889-017-4533-z] [Medline: 28683797]
7. Vardavas CI, Anagnostopoulos N, Kougias M, Evangelopoulou V, Connolly GN, Behrakis PK. Short-term pulmonary effects of using an electronic cigarette: impact on respiratory flow resistance, impedance, and exhaled nitric oxide. Chest 2012 Jun;141(6):1400-1406. [doi: 10.1378/chest.11-2443] [Medline: 22194587]
8. Moheimani RS, Bhetraratana M, Yin F, Peters KM, Gornbein J, Araujo JA, et al. Increased cardiac sympathetic activity and oxidative stress in habitual electronic cigarette users: implications for cardiovascular risk. JAMA Cardiol 2017 Mar 1;2(3):278-284 [FREE Full text] [doi: 10.1001/jamacardio.2016.5303] [Medline: 28146259]
9. Bahl V, Lin S, Xu N, Davis B, Wang Y, Talbot P. Comparison of electronic cigarette refill fluid cytotoxicity using embryonic and adult models. Reprod Toxicol 2012 Dec;34(4):529-537. [doi: 10.1016/j.reprotox.2012.08.001] [Medline: 22989551]
10. Behar RZ, Davis B, Wang Y, Bahl V, Lin S, Talbot P. Identification of toxicants in cinnamon-flavored electronic cigarette refill fluids. Toxicol In Vitro 2014 Mar;28(2):198-208. [doi: 10.1016/j.tiv.2013.10.006] [Medline: 24516877]
11. Behar RZ, Davis B, Bahl V, Lin S, Talbot P. Commentary in response to the letter from Farsalinos et al. regarding our publication entitled: 'identification of toxicants in cinnamon-flavored electronic cigarette refill fluids'. Toxicol In Vitro 2014 Dec;28(8):1521-1522. [doi: 10.1016/j.tiv.2014.07.001] [Medline: 25017651]
12. Yu V, Rahimy M, Korrapati A, Xuan Y, Zou AE, Krishnan AR, et al. Electronic cigarettes induce DNA strand breaks and cell death independently of nicotine in cell lines. Oral Oncol 2016 Jan;52:58-65 [FREE Full text] [doi: 10.1016/j.oraloncology.2015.10.018] [Medline: 26547127]

13. Lerner CA, Sundar IK, Yao H, Gerloff J, Ossip DJ, McIntosh S, et al. Vapors produced by electronic cigarettes and e-juices with flavorings induce toxicity, oxidative stress, and inflammatory response in lung epithelial cells and in mouse lung. PLoS One 2015;10(2):e0116732 [FREE Full text] [doi: 10.1371/journal.pone.0116732] [Medline: 25658421]
14. Sussan TE, Gajghate S, Thimmulappa RK, Ma J, Kim J, Sudini K, et al. Exposure to electronic cigarettes impairs pulmonary anti-bacterial and anti-viral defenses in a mouse model. PLoS One 2015;10(2):e0116861 [FREE Full text] [doi: 10.1371/journal.pone.0116861] [Medline: 25651083]
15. Schweitzer KS, Chen SX, Law S, Van Demark M, Poirier C, Justice MJ, et al. Endothelial disruptive proinflammatory effects of nicotine and e-cigarette vapor exposures. Am J Physiol Lung Cell Mol Physiol 2015 Jul 15;309(2):L175-L187 [FREE Full text] [doi: 10.1152/ajplung.00411.2014] [Medline: 25979079]
16. Drummond MB, Upson D. Electronic cigarettes. Potential harms and benefits. Ann Am Thorac Soc 2014 Feb;11(2):236-242 [FREE Full text] [doi: 10.1513/AnnalsATS.201311-391FR] [Medline: 24575993]
17. Hua M, Talbot P. Potential health effects of electronic cigarettes: a systematic review of case reports. Prev Med Rep 2016 Dec;4:169-178 [FREE Full text] [doi: 10.1016/j.pmedr.2016.06.002] [Medline: 27413679]
18. Pisinger C, Døssing M. A systematic review of health effects of electronic cigarettes. Prev Med 2014 Dec;69:248-260 [FREE Full text] [doi: 10.1016/j.ypmed.2014.10.009] [Medline: 25456810]
19. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
20. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
21. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
22. Hua M, Yip H, Talbot P. Mining data on usage of electronic nicotine delivery systems (ENDS) from YouTube videos. Tob Control 2013 Mar;22(2):103-106. [doi: 10.1136/tobaccocontrol-2011-050226] [Medline: 22116832]
23. Behar RZ, Hua M, Talbot P. Puffing topography and nicotine intake of electronic cigarette users. PLoS One 2015;10(2):e0117222 [FREE Full text] [doi: 10.1371/journal.pone.0117222] [Medline: 25664463]
24. Hedley J. jsoup Java HTML Parser, with best of DOM, CSS, and jquery. 2009. URL: https://jsoup.org/ [accessed 2019-02-07]
25. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
26. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
27. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. J Biomed Inform 2003 Dec;36(6):414-432 [FREE Full text] [doi: 10.1016/j.jbi.2003.11.002] [Medline: 14759816]
28. Denecke K, Soltani N. Where Humans Meet Machines: Innovative Solutions for Knotty Natural-Language Problems. New York: Springer; 2013.
29. Freelon D. dfreelon. 2017. ReCal: reliability calculation for the masses. URL: http://dfreelon.org/utils/recalfront/ [accessed 2017-11-02]
30. Frank E, Hall MA, Witten IH. The WEKA Workbench in Data Mining: Practical Machine Learning Tools and Techniques. Fourth Edition. San Francisco: Morgan Kaufmann Publishers; 2016.
31. Sun Y, Wong AK, Kamel MS. Classification of imbalanced data: a review. Int J Patt Recogn Artif Intell 2011;23(4):687-719. [doi: 10.1142/s0218001409007326]
32. Vannier S, Ronziere T, Ferre JC, Lassalle V, Verin M. Reversible cerebral vasoconstriction syndrome triggered by an electronic cigarette: case report. Eur J Neurol 2015 May;22(5):e64-e65. [doi: 10.1111/ene.12657] [Medline: 25846567]
33. McCauley L, Markin C, Hosmer D. An unexpected consequence of electronic cigarette use. Chest 2012 Apr;141(4):1110-1113. [doi: 10.1378/chest.11-1334] [Medline: 22474155]
34. Khan MS, Khateeb F, Akhtar J, Khan Z, Lal A, Kholodovych V, et al. Organizing pneumonia related to electronic cigarette use: a case report and review of literature. Clin Respir J 2018 Mar;12(3):1295-1299. [doi: 10.1111/crj.12775] [Medline: 29392888]
35. Viswam D, Trotter S, Burge PS, Walters GI. Respiratory failure caused by lipoid pneumonia from vaping e-cigarettes. BMJ Case Rep 2018 Jul 6;2018:pii: bcr-2018-224350. [doi: 10.1136/bcr-2018-224350] [Medline: 29982176]
36. Dawkins L, Turner J, Crowe E. Nicotine derived from the electronic cigarette improves time-based prospective memory in abstinent smokers. Psychopharmacology (Berl) 2013 Jun;227(3):377-384. [doi: 10.1007/s00213-013-2983-2] [Medline: 23344557]
37. Dawkins L, Turner J, Hasna S, Soar K. The electronic-cigarette: effects on desire to smoke, withdrawal symptoms and cognition. Addict Behav 2012 Aug;37(8):970-973. [doi: 10.1016/j.addbeh.2012.03.004] [Medline: 22503574]
38. Berg CJ, Barr DB, Stratton E, Escoffery C, Kegler M. Attitudes toward e-cigarettes, reasons for initiating e-cigarette use, and changes in smoking behavior after initiation: a pilot longitudinal study of regular cigarette smokers. Open J Prev Med 2014 Oct;4(10):789-800 [FREE Full text] [doi: 10.4236/ojpm.2014.410089] [Medline: 25621193]

39. Motooka Y, Matsui T, Slaton RM, Umetsu R, Fukuda A, Naganuma M, et al. Adverse events of smoking cessation treatments (nicotine replacement therapy and non-nicotine prescription medication) and electronic cigarettes in the Food and Drug Administration Adverse Event Reporting System, 2004-2016. SAGE Open Med 2018;6:2050312118777953 [FREE Full text] [doi: 10.1177/2050312118777953] [Medline: 29844912]
40. Food and Drug Administration. 2019. Some E-cigarette Users Are Having Seizures, Most Reports Involving Youth and Young Adults. URL: https://www.fda.gov/tobacco-products/ctp-newsroom/some-e-cigarette-users-are-having-seizures-most-reports-involving-youth-and-young-adults [accessed 2019-11-20]
41. Meo SA, Al Asiri SA. Effects of electronic cigarette smoking on human health. Eur Rev Med Pharmacol Sci 2014;18(21):3315-3319 [FREE Full text] [Medline: 25487945]
42. Zvolensky MJ, Garey L, Mayorga NA, Peraza N, Rogers AH, Ditre JW, et al. Pain severity and anxiety sensitivity interplay among exclusive and dual electronic cigarette users. J Stud Alcohol Drugs 2019 Mar;80(2):211-219. [doi: 10.15288/jsad.2019.80.211] [Medline: 31014466]
43. Chen IL. FDA summary of adverse events on electronic cigarettes. Nicotine Tob Res 2013 Feb;15(2):615-616. [doi: 10.1093/ntr/nts145] [Medline: 22855883]
44. Baweja R, Curci KM, Yingst J, Veldheer S, Hrabovsky S, Wilson SJ, et al. Views of experienced electronic cigarette users. Addict Res Theory 2016;24(1):80-88 [FREE Full text] [doi: 10.3109/16066359.2015.1077947] [Medline: 29176939]
45. Volesky KD, Maki A, Scherf C, Watson LM, Cassol E, Villeneuve PJ. Characteristics of e-cigarette users and their perceptions of the benefits, harms and risks of e-cigarette use: survey results from a convenience sample in Ottawa, Canada. Health Promot Chronic Dis Prev Can 2016 Jul;36(7):130-138 [FREE Full text] [doi: 10.24095/hpcdp.36.7.02] [Medline: 27409988]
46. Mohamed MH, Rahman A, Jamshed S, Mahmood S. Effectiveness and safety of electronic cigarettes among sole and dual user vapers in Kuantan and Pekan, Malaysia: a six-month observational study. BMC Public Health 2018 Aug 20;18(1):1028 [FREE Full text] [doi: 10.1186/s12889-018-5951-2] [Medline: 30126382]
47. Caponnetto P, Campagna D, Cibella F, Morjaria JB, Caruso M, Russo C, et al. EffiCiency and Safety of an eLectronic cigAreTte (ECLAT) as tobacco cigarettes substitute: a prospective 12-month randomized control design study. PLoS One 2013;8(6):e66317 [FREE Full text] [doi: 10.1371/journal.pone.0066317] [Medline: 23826093]
48. Cravo AS, Bush J, Sharma G, Savioz R, Martin C, Craige S, et al. A randomised, parallel group study to evaluate the safety profile of an electronic vapour product over 12 weeks. Regul Toxicol Pharmacol 2016 Nov 15;81(Suppl 1):S1-14 [FREE Full text] [doi: 10.1016/j.yrtph.2016.10.003] [Medline: 27769828]
49. Cooper M, Harrell MB, Perry CL. A qualitative approach to understanding real-world electronic cigarette use: implications for measurement and regulation. Prev Chronic Dis 2016 Jan 14;13:E07 [FREE Full text] [doi: 10.5888/pcd13.150502] [Medline: 26766848]
50. Csere A, Sahai-Srivastava S. The Development of New Daily Persistent Headache after Electronic Cigarette Use in an Adolescent. In: Proceedings of the American Headache Society 57th Annual Scientific Meeting. 2015 Presented at: AHS'15; June 18-21, 2015; Washington, DC.
51. OECD SIDS. OECD's Work on Co-operating in the Investigation. 2002. Screening information data set (SIDS): SIDS initial assessment report for SIAM 14; glycerol: CAS no: 56-81-5. URL: https://hpvchemicals.oecd.org/ui/handler.axd?id=4b0a2d87-3183-40d4-84f5-0e118c647b19 [accessed 2018-06-02]
52. Cosmetic Ingredient Review. 2014. Safety Assessment of Glycerin as Used in Cosmetics: Draft Report for Panel Review. URL: https://www.cir-safety.org/sites/default/files/glycerin.pdf [accessed 2018-02-01]
53. Venable JR, McClimans CD, Flake RE, Dimick DB. A fertility study of male employees engaged in the manufacture of glycerine. J Occup Med 1980 Feb;22(2):87-91. [doi: 10.1097/00043764-198002000-00005] [Medline: 7373447]
54. National Academies of Sciences, Engineering, and Medicine. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.
55. Alzahrani T, Pena I, Temesgen N, Glantz SA. Association between electronic cigarette use and myocardial infarction. Am J Prev Med 2018 Oct;55(4):455-461 [FREE Full text] [doi: 10.1016/j.amepre.2018.05.004] [Medline: 30166079]
56. Li D, Sundar IK, McIntosh S, Ossip DJ, Goniewicz ML, O'Connor RJ, et al. Association of smoking and electronic cigarette use with wheezing and related respiratory symptoms in adults: cross-sectional results from the Population Assessment of Tobacco and Health (PATH) study, wave 2. Tob Control 2019 Feb 13. [doi: 10.1136/tobaccocontrol-2018-054694] [Medline: 30760629]
57. McConnell R, Barrington-Trimis JL, Wang K, Urman R, Hong H, Unger J, et al. Electronic cigarette use and respiratory symptoms in adolescents. Am J Respir Crit Care Med 2017 Apr 15;195(8):1043-1049 [FREE Full text] [doi: 10.1164/rccm.201604-0804OC] [Medline: 27806211]
58. Hedman L, Backman H, Stridsman C, Bosson JA, Lundbäck M, Lindberg A, et al. Association of electronic cigarette use with smoking habits, demographic factors, and respiratory symptoms. JAMA Netw Open 2018 Jul 6;1(3):e180789 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0789] [Medline: 30646032]
59. Moazed F, Calfee CS. The canary in the coal mine is coughing: electronic cigarettes and respiratory symptoms in adolescents. Am J Respir Crit Care Med 2017 Apr 15;195(8):974-976 [FREE Full text] [doi: 10.1164/rccm.201611-2259ED] [Medline: 28409686]

60. Agustin M, Yamamoto M, Cabrera F, Eusebio R. Diffuse alveolar hemorrhage induced by vaping. Case Rep Pulmonol 2018;2018:9724530 [FREE Full text] [doi: 10.1155/2018/9724530] [Medline: 29984031]
61. Shields PG, Berman M, Brasky TM, Freudenheim JL, Mathe E, McElroy JP, et al. A review of pulmonary toxicity of electronic cigarettes in the context of smoking: a focus on inflammation. Cancer Epidemiol Biomarkers Prev 2017 Aug;26(8):1175-1191 [FREE Full text] [doi: 10.1158/1055-9965.EPI-17-0358] [Medline: 28642230]
62. Tang L. Pain Associated with the Use of Electronic Cigarettes. IntechOpen 2019 [FREE Full text] [doi: 10.5772/intechopen.85481]
63. Kivrak T, Sunbul M, Durmus E, Dervisova R, Sari I, Yesildag O. Acute myocardial infarction due to liquid nicotine in a young man. Ther Adv Cardiovasc Dis 2014 Feb;8(1):32-34. [doi: 10.1177/1753944713515765] [Medline: 24492987]
64. Javed F, Kellesarian SV, Sundar IK, Romanos GE, Rahman I. Recent updates on electronic cigarette aerosol and inhaled nicotine effects on periodontal and pulmonary tissues. Oral Dis 2017 Nov;23(8):1052-1057 [FREE Full text] [doi: 10.1111/odi.12652] [Medline: 28168771]
65. Kaur G, Pinkston R, Mclemore B, Dorsey WC, Batra S. Immunological and toxicological risk assessment of e-cigarettes. Eur Respir Rev 2018 Mar 31;27(147):pii: 170119 [FREE Full text] [doi: 10.1183/16000617.0119-2017] [Medline: 29491036]
66. Anderson C, Majeste A, Hanus J, Wang S. E-cigarette aerosol exposure induces reactive oxygen species, DNA damage, and cell death in vascular endothelial cells. Toxicol Sci 2016 Dec;154(2):332-340. [doi: 10.1093/toxsci/kfw166] [Medline: 27613717]
67. Muthumalage T, Prinz M, Ansah KO, Gerloff J, Sundar IK, Rahman I. Inflammatory and oxidative responses induced by exposure to commonly used e-cigarette flavoring chemicals and flavored e-liquids without nicotine. Front Physiol 2017;8:1130 [FREE Full text] [doi: 10.3389/fphys.2017.01130] [Medline: 29375399]
68. Hallagan JB, Hall RL. Under the conditions of intended use - New developments in the FEMA GRAS program and the safety assessment of flavor ingredients. Food Chem Toxicol 2009 Feb;47(2):267-278. [doi: 10.1016/j.fct.2008.11.011] [Medline: 19041920]
69. Behar RZ, Wang Y, Talbot P. Comparing the cytotoxicity of electronic cigarette fluids, aerosols and solvents. Tob Control 2018 May;27(3):325-333 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053472] [Medline: 28596276]
70. Behar RZ, Luo W, McWhirter KJ, Pankow JF, Talbot P. Analytical and toxicological evaluation of flavor chemicals in electronic cigarette refill fluids. Sci Rep 2018 May 29;8(1):8288 [FREE Full text] [doi: 10.1038/s41598-018-25575-6] [Medline: 29844439]
71. Omaiye EE, McWhirter KJ, Luo W, Tierney PA, Pankow JF, Talbot P. High concentrations of flavor chemicals are present in electronic cigarette refill fluids. Sci Rep 2019 Feb 21;9(1):2468 [FREE Full text] [doi: 10.1038/s41598-019-39550-2] [Medline: 30792477]
72. Hua M, Omaiye EE, Luo W, McWhirter KJ, Pankow JF, Talbot P. Identification of cytotoxic flavor chemicals in top-selling electronic cigarette refill fluids. Sci Rep 2019 Feb 26;9(1):2782 [FREE Full text] [doi: 10.1038/s41598-019-38978-w] [Medline: 30808901]
73. Farsalinos KE, Kistler KA, Gillman G, Voudris V. Evaluation of electronic cigarette liquids and aerosol for the presence of selected inhalation toxins. Nicotine Tob Res 2015 Feb;17(2):168-174 [FREE Full text] [doi: 10.1093/ntr/ntu176] [Medline: 25180080]
74. Varlet V, Farsalinos K, Augsburger M, Thomas A, Etter J. Toxicity assessment of refill liquids for electronic cigarettes. Int J Environ Res Public Health 2015 Apr 30;12(5):4796-4815 [FREE Full text] [doi: 10.3390/ijerph120504796] [Medline: 25941845]
75. Allen JG, Flanigan SS, LeBlanc M, Vallarino J, MacNaughton P, Stewart JH, et al. Flavoring chemicals in e-cigarettes: diacetyl, 2, 3-pentanedione, and acetoin in a sample of 51 products, including fruit-, candy-, and cocktail-flavored e-cigarettes. Environ Health Perspect 2016 Jun;124(6):733-739 [FREE Full text] [doi: 10.1289/ehp.1510185] [Medline: 26642857]
76. Clapp PW, Pawlak EA, Lackey JT, Keating JE, Reeber SL, Glish GL, et al. Flavored e-cigarette liquids and cinnamaldehyde impair respiratory innate immune cell function. Am J Physiol Lung Cell Mol Physiol 2017 Aug 1;313(2):L278-L292 [FREE Full text] [doi: 10.1152/ajplung.00452.2016] [Medline: 28495856]
77. Clapp PW, Jaspers I. Electronic cigarettes: their constituents and potential links to asthma. Curr Allergy Asthma Rep 2017 Oct 5;17(11):79 [FREE Full text] [doi: 10.1007/s11882-017-0747-5] [Medline: 28983782]
78. Clapp PW, Lavrich KS, van Heusden CA, Lazarowski ER, Carson JL, Jaspers I. Cinnamaldehyde in flavored e-cigarette liquids temporarily suppresses bronchial epithelial cell ciliary motility by dysregulation of mitochondrial function. Am J Physiol Lung Cell Mol Physiol 2019 Mar 1;316(3):L470-L486. [doi: 10.1152/ajplung.00304.2018] [Medline: 30604630]
79. Ni Y, Zhang G, Kokot S. Simultaneous spectrophotometric determination of maltol, ethyl maltol, vanillin and ethyl vanillin in foods by multivariate calibration and artificial neural networks. Food Chem 2005;89(3):465-473. [doi: 10.1016/j.foodchem.2004.05.037]
80. Yue Y, Liu J, Yao M, Yao X, Fan J, Ji H. The investigation of the binding behavior between ethyl maltol and human serum albumin by multi-spectroscopic methods and molecular docking. Spectrochim Acta A Mol Biomol Spectrosc 2012 Oct;96:316-323. [doi: 10.1016/j.saa.2012.05.041] [Medline: 22705675]
81. Gonullu H, Gonullu E, Karadas S, Arslan M, Kalemci O, Aycan A, et al. The levels of trace elements and heavy metals in patients with acute migraine headache. J Pak Med Assoc 2015 Jul;65(7):694-697 [FREE Full text] [Medline: 26160074]

82. Agency for Toxic Substances and Disease Registry (ATSDR). Atlanta, GA: US Department of Health and Human Services, Public Health Service; 2004 Sep. Toxicological profile for Copper. URL: https://www.atsdr.cdc.gov/toxprofiles/tp132.pdf [accessed 2019-07-25]
83. Brodkin E, Copes R, Mattman A, Kennedy J, Kling R, Yassi A. Lead and mercury exposures: interpretation and action. Can Med Assoc J 2007 Jan 2;176(1):59-63 [FREE Full text] [doi: 10.1503/cmaj.060790] [Medline: 17200393]
84. OECD's Work on Co-operating in the Investigation. 2001. Screening information data set (SIDS): SIDS initial assessment profile for SIAM 11; Dipropylene glycol (mixed isomers and dominant isomer). URL: https://hpvchemicals.oecd.org/ui/handler.axd?id=40da06b1-a855-4c0c-bc21-bbc856dca725 [accessed 2019-11-20]
85. NJ Health. 2009. Hazardous Substances Fact Sheet: Propylene Glycol. URL: http://nj.gov/health/eoh/rtkweb/documents/fs/3595.pdf [accessed 2019-07-25]
86. Centers for Disease Control and Prevention. 2018. NIOSH: Nicotine. URL: https://www.cdc.gov/niosh/npg/npgd0446.html
87. Mishra A, Chaturvedi P, Datta S, Sinukumar S, Joshi P, Garg A. Harmful effects of nicotine. Indian J Med Paediatr Oncol 2015;36(1):24-31 [FREE Full text] [doi: 10.4103/0971-5851.151771] [Medline: 25810571]
88. Taylor FR. Tobacco, nicotine, and headache. Headache 2015;55(7):1028-1044. [doi: 10.1111/head.12620] [Medline: 26140522]
89. Ash M, Ash I. Handbook of Green Chemicals. Second Edition. New York, NY: Synapse Information Resources Inc; 2006.
90. Assi MA, Hezmee MN, Haron AW, Sabri MY, Rajion MA. The detrimental effects of lead on human and animal health. Vet World 2016 Jun;9(6):660-671 [FREE Full text] [doi: 10.14202/vetworld.2016.660-671] [Medline: 27397992]
91. NJ Health. 2009. Hazardous Substances Fact Sheet: Benzaldehyde. URL: https://nj.gov/health/eoh/rtkweb/documents/fs/0196.pdf [accessed 2019-07-25]
92. Environmental XPRT. 2010. Overview of Airborne metals regulations, exposure limits, health effects, and contemporary research case study. URL: https://www.environmental-expert.com/articles/overview-of-airborne-metals-regulations-exposure-limits-health-effects-and-contemporary-research-cas-450906 [accessed 2019-07-25]
93. Jaishankar M, Tseten T, Anbalagan N, Mathew BB, Beeregowda KN. Toxicity, mechanism and health effects of some heavy metals. Interdiscip Toxicol 2014 Jun;7(2):60-72 [FREE Full text] [doi: 10.2478/intox-2014-0009] [Medline: 26109881]
94. Centers for Disease Control and Prevention. 2018. NIOSH: Nicotine: Systemic Agent. URL: https://www.cdc.gov/niosh/ershdb/emergencyresponsecard_29750028.html [accessed 2019-07-25]
95. Agency for Toxic Substances and Disease Registry. 2014. Medical Management Guidelines for Phenol. URL: https://www.atsdr.cdc.gov/mmg/mmg.asp?id=144&tid=27#bookmark05 [accessed 2019-07-25]
96. Callahan-Lyon P. Electronic cigarettes: human health effects. Tob Control 2014 May;23(Suppl 2):ii36-ii40 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051470] [Medline: 24732161]
97. Pesta DH, Angadi SS, Burtscher M, Roberts CK. The effects of caffeine, nicotine, ethanol, and tetrahydrocannabinol on exercise performance. Nutr Metab (Lond) 2013 Dec 13;10(1):71 [FREE Full text] [doi: 10.1186/1743-7075-10-71] [Medline: 24330705]
98. NJ Health. 1998. Hazardous Substances Fact Sheet: ethyl butyrate. URL: http://nj.gov/health/eoh/rtkweb/documents/fs/0862.pdf [accessed 2019-11-20]
99. Park RM, Gilbert SJ. Pulmonary impairment and risk assessment in a diacetyl-exposed population: microwave popcorn workers. J Occup Environ Med 2018 Jun;60(6):496-506 [FREE Full text] [doi: 10.1097/JOM.0000000000001303] [Medline: 29443707]
100. Toxnet. 2015. Triacetin. URL: https://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+585 [accessed 2019-07-25]
101. Wieslander G, Norbäck D, Lindgren T. Experimental exposure to propylene glycol mist in aviation emergency training: acute ocular and respiratory effects. Occup Environ Med 2001 Oct;58(10):649-655 [FREE Full text] [doi: 10.1136/oem.58.10.649] [Medline: 11555686]
102. Dicpinigaitis PV. Effect of tobacco and electronic cigarette use on cough reflex sensitivity. Pulm Pharmacol Ther 2017 Dec;47:45-48. [doi: 10.1016/j.pupt.2017.01.013] [Medline: 28185897]
103. Nemery B. Metal toxicity and the respiratory tract. Eur Respir J 1990 Feb;3(2):202-219. [Medline: 2178966]
104. Boulay ME, Henry C, Bossé Y, Boulet LP, Morissette MC. Acute effects of nicotine-free and flavour-free electronic cigarette use on lung functions in healthy and asthmatic individuals. Respir Res 2017 Feb 10;18(1):33 [FREE Full text] [doi: 10.1186/s12931-017-0518-9] [Medline: 28183298]
105. Cooper RG. Zinc toxicology following particulate inhalation. Indian J Occup Environ Med 2008 Apr;12(1):10-13 [FREE Full text] [doi: 10.4103/0019-5278.40809] [Medline: 20040991]
106. United States Environmental Protection Agency. 2000. Selenium Compounds. URL: https://www.epa.gov/sites/production/files/2016-09/documents/selenium-compounds.pdf [accessed 2019-07-25]
107. Norbäck D, Wieslander G, Edling C. Occupational exposure to volatile organic compounds (VOCs), and other air pollutants from the indoor application of water-based paints. Ann Occ Hyg 1995;39(6):783-794. [doi: 10.1016/0003-4878(95)00034-8]
108. Paik JH, Kang S, Durey A, Kim JH, Kim AJ. Symptomatic bradycardia due to nicotine intoxication. Rev Bras Ter Intensiva 2018 Mar;30(1):121-126 [FREE Full text] [doi: 10.5935/0103-507x.20180018] [Medline: 29742227]

109. Quirce S, Fernández-Nieto M, del Pozo V, Sastre B, Sastre J. Occupational asthma and rhinitis caused by eugenol in a hairdresser. Allergy 2008 Jan;63(1):137-138. [doi: 10.1111/j.1398-9995.2007.01525.x] [Medline: 17711554]
110. Bein K, Leikauf GD. Acrolein - a pulmonary hazard. Mol Nutr Food Res 2011 Sep;55(9):1342-1360. [doi: 10.1002/mnfr.201100279] [Medline: 21994168]
111. Centers for Disease Control and Prevention. 2017. Flavorings-related lung disease. URL: https://www.cdc.gov/niosh/topics/flavorings/exposure.html [accessed 2019-07-25]
112. Sisson JH. Alcohol and airways function in health and disease. Alcohol 2007 Aug;41(5):293-307 [FREE Full text] [doi: 10.1016/j.alcohol.2007.06.003] [Medline: 17764883]
113. Gorguner M, Akgun M. Acute inhalation injury. Eurasian J Med 2010 Apr;42(1):28-35 [FREE Full text] [doi: 10.5152/eajm.2010.09] [Medline: 25610115]
114. Chang CC, Incaudo GA, Gershwin ME. Diseases of the Sinuses: A Comprehensive Textbook of Diagnosis and Treatment. New York, NY: Springer-Verlag; 2014.
115. Sunderman FW. Nasal toxicity, carcinogenicity, and olfactory uptake of metals. Ann Clin Lab Sci 2001 Jan;31(1):3-24. [Medline: 11314863]
116. Kim SA, Smith S, Beauchamp C, Song Y, Chiang M, Giuseppetti A, et al. Cariogenic potential of sweet flavors in electronic-cigarette liquids. PLoS One 2018;13(9):e0203717 [FREE Full text] [doi: 10.1371/journal.pone.0203717] [Medline: 30192874]
117. Wychowanski P, Malkiewicz K. Evaluation of metal ion concentration in hard tissues of teeth in residents of Central Poland. Biomed Res Int 2017;2017:6419709 [FREE Full text] [doi: 10.1155/2017/6419709] [Medline: 28197416]
118. Liu S, Wu T, Zhou X, Zhang B, Huo S, Yang Y, et al. Nicotine is a risk factor for dental caries: an study. J Dent Sci 2018 Mar;13(1):30-36 [FREE Full text] [doi: 10.1016/j.jds.2017.09.006] [Medline: 30895091]
119. Maridet C, Atge B, Amici J, Taïeb A, Milpied B. The electronic cigarette: the new source of nickel contact allergy of the 21st century? Contact Dermatitis 2015 Jul;73(1):49-50. [doi: 10.1111/cod.12373] [Medline: 25801540]
120. Shim TN, Kosztyuova T. Allergic contact dermatitis to electronic cigarette. Dermatitis 2018;29(2):94-95. [doi: 10.1097/DER.0000000000000337] [Medline: 29494384]
121. Nguyen TT, Higashi T, Kambayashi Y, Anyenda EO, Michigami Y, Hara J, et al. A longitudinal study of association between heavy metals and itchy eyes, coughing in chronic cough patients: related with non-immunoglobulin E mediated mechanism. Int J Environ Res Public Health 2016 Jan 7;13(1):pii: E110 [FREE Full text] [doi: 10.3390/ijerph13010110] [Medline: 26751467]
122. Andersen KE, Storrs FJ. [Skin irritation caused by propylene glycols]. Hautarzt 1982 Jan;33(1):12-14. [Medline: 7085276]
123. Lee IW, Ahn SK, Choi EH, Lee SH. Urticarial reaction following the inhalation of nicotine in tobacco smoke. Br J Dermatol 1998 Mar;138(3):486-488. [doi: 10.1046/j.1365-2133.1998.02130.x] [Medline: 9580805]
124. Agency for Toxic Substances and Disease Registry. Atlanta, GA: US Department of Health and Human Services, Public Health Service; 2015. Public Health Statement for Propylene Glycol. URL: https://www.atsdr.cdc.gov/toxprofiles/tp189-c1-b.pdf [accessed 2019-11-20]
125. Sarici G, Cinar S, Armutcu F, Altinyazar C, Koca R, Tekin NS. Oxidative stress in acne vulgaris. J Eur Acad Dermatol Venereol 2010 Jul;24(7):763-767. [doi: 10.1111/j.1468-3083.2009.03505.x] [Medline: 19943837]
126. Mills CM, Peters TJ, Finlay AY. Does smoking influence acne? Clin Exp Dermatol 1993 Mar;18(2):100-101. [doi: 10.1111/j.1365-2230.1993.tb00986.x] [Medline: 8481981]
127. Misery L. Nicotine effects on skin: are they positive or negative? Exp Dermatol 2004 Nov;13(11):665-670. [doi: 10.1111/j.0906-6705.2004.00274.x] [Medline: 15500638]
128. Wait M. WebMD. 2014. Why Tobacco Use May Make Heartburn Worse. URL: https://www.webmd.com/heartburn-gerd/features/heartburn-tobacco-connection [accessed 2019-07-25]
129. Mueller PD, Benowitz NL. Toxicologic causes of acute abdominal disorders. Emerg Med Clin North Am 1989 Aug;7(3):667-682. [Medline: 2663462]
130. Williams M, Villarreal A, Bozhilov K, Lin S, Talbot P. Metal and silicate particles including nanoparticles are present in electronic cigarette cartomizer fluid and aerosol. PLoS One 2013;8(3):e57987 [FREE Full text] [doi: 10.1371/journal.pone.0057987] [Medline: 23526962]
131. Olmedo P, Goessler W, Tanda S, Grau-Perez M, Jarmul S, Aherrera A, et al. Metal concentrations in e-cigarette liquid and aerosol samples: the contribution of metallic coils. Environ Health Perspect 2018 Feb 21;126(2):027010 [FREE Full text] [doi: 10.1289/EHP2175] [Medline: 29467105]
132. Geiger A, Cooper J. Overview of Airborne Metals: Regulations, Exposure Limits, Health Effects, and Contemporary Research. Portland, OR: Cooper Environmental Services; 2010.
133. Aherrera A, Olmedo P, Grau-Perez M, Tanda S, Goessler W, Jarmul S, et al. The association of e-cigarette use with exposure to nickel and chromium: a preliminary study of non-invasive biomarkers. Environ Res 2017 Nov;159:313-320. [doi: 10.1016/j.envres.2017.08.014] [Medline: 28837903]
134. Sato T, Shimosato T, Klinman DM. Silicosis and lung cancer: current perspectives. Lung Cancer (Auckl) 2018;9:91-101 [FREE Full text] [doi: 10.2147/LCTT.S156376] [Medline: 30498384]

135. Palazzolo DL. Electronic cigarettes and vaping: a new challenge in clinical medicine and public health. A literature review. Front Public Health 2013 Nov 18;1:56 [FREE Full text] [doi: 10.3389/fpubh.2013.00056] [Medline: 24350225]
136. Hess CA, Olmedo P, Navas-Acien A, Goessler W, Cohen JE, Rule AM. E-cigarettes as a source of toxic and potentially carcinogenic metals. Environ Res 2017 Jan;152:221-225 [FREE Full text] [doi: 10.1016/j.envres.2016.09.026] [Medline: 27810679]
137. Dalton P, Soreth B, Maute C, Novaleski C, Banton M. Lack of respiratory and ocular effects following acute propylene glycol exposure in healthy humans. Inhal Toxicol 2018 Feb;30(3):124-132. [doi: 10.1080/08958378.2018.1470207] [Medline: 29764241]
138. Jensen RP, Strongin RM, Peyton DH. Solvent chemistry in the electronic cigarette reaction vessel. Sci Rep 2017 Feb 14;7:42549 [FREE Full text] [doi: 10.1038/srep42549] [Medline: 28195231]
139. Occupational Safety and Health Administration. 1910.1048 App C - Medical surveillance - Formaldehyde. URL: https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.1048AppC [accessed 2019-11-20]
140. Cotta KI, Stephen CD, Mohammad NU. A review on the safety of inhalation of propylene glycol in e-cigarettes. Glob J Pharm Pharm Sci 2017;2(2):1-2. [doi: 10.19080/gjpps.2017.02.555584]
141. Watson AY, Bates RR, Kennedy D. Health effects of aldehydes and alcohols in mobile source emissions. In: Air Pollution, the Automobile, and Public Health. Washington, DC: National Academies Press; 1988.
142. Kawanishi M, Matsuda T, Yagi T. Genotoxicity of formaldehyde: molecular basis of DNA damage and mutation. Front Environ Sci 2014;2:36. [doi: 10.3389/fenvs.2014.00036]
143. Drugs. 2019. Nicotine Side Effects. URL: https://www.drugs.com/sfx/nicotine-side-effects.html#moreResources [accessed 2019-07-17]
144. McLaughlin I, Dani JA, de Biasi M. Nicotine withdrawal. Curr Top Behav Neurosci 2015;24:99-123 [FREE Full text] [doi: 10.1007/978-3-319-13482-6_4] [Medline: 25638335]
145. Duell AK, Pankow JF, Peyton DH. Free-Base Nicotine Determination in Electronic Cigarette Liquids by 1H NMR Spectroscopy. Chem Res Toxicol 2018 Jun 18;31(6):431-434 [FREE Full text] [doi: 10.1021/acs.chemrestox.8b00097] [Medline: 29775302]
146. Henningfield J, Pankow J, Garrett B. Ammonia and other chemical base tobacco additives and cigarette nicotine delivery: issues and research needs. Nicotine Tob Res 2004 Apr;6(2):199-205. [doi: 10.1080/14622200420002024 72] [Medline: 15203793]
147. Center for Disease Control and Prevention. 2019 Nov 10. Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2019-11-10]
148. Layden JE, Ghinai I, Pray I, Kimball A, Layer M, Tenforde M, et al. Pulmonary Illness Related to E-Cigarette Use in Illinois and Wisconsin - Preliminary Report. N Engl J Med 2019 Sep 6. [doi: 10.1056/NEJMoa1911614] [Medline: 31491072]
149. Davidson K, Brancato A, Heetderks P, Mansour W, Matheis E, Nario M, et al. Outbreak of Electronic-Cigarette-Associated Acute Lipoid Pneumonia - North Carolina, July-August 2019. MMWR Morb Mortal Wkly Rep 2019 Sep 13;68(36):784-786 [FREE Full text] [doi: 10.15585/mmwr.mm6836e1] [Medline: 31513559]
150. Henry TS, Kanne JP, Kligerman SJ. Imaging of vaping-associated lung disease. N Engl J Med 2019 Oct 10;381(15):1486-1487. [doi: 10.1056/NEJMc1911995] [Medline: 31491070]

Abbreviations

CDC: Centers for Disease Control and Prevention
COPD: chronic obstructive pulmonary disorder
CTP: Center for Tobacco Products
EC: electronic cigarette
FDA: Food and Drug Administration
PATH: Population Assessment of Tobacco Health
UMLS: Unified Medical Language System
VAPI: vaping-associated pulmonary illness

Edited by G Eysenbach; submitted 29.07.19; peer-reviewed by L Wang, K McCausland; comments to author 03.10.19; revised version received 11.10.19; accepted 20.10.19; published 03.01.20.

Please cite as:
Hua M, Sadah S, Hristidis V, Talbot P
Health Effects Associated With Electronic Cigarette Use: Automated Mining of Online Forums
J Med Internet Res 2020;22(1):e15684
URL: https://www.jmir.org/2020/1/e15684
doi:10.2196/15684
PMID:31899452

Original Paper

Lifestyle Disease Surveillance Using Population Search Behavior: Feasibility Study

Shahan Ali Memon[1], MSc; Saquib Razak[2], PhD; Ingmar Weber[3], PhD

[1]Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

[2]Carnegie Mellon University, Doha, Qatar

[3]Social Computing Department, Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Shahan Ali Memon, MSc
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
United States
Phone: 1 4123785380
Email: samemon@cs.cmu.edu

Abstract

Background: As the process of producing official health statistics for lifestyle diseases is slow, researchers have explored using Web search data as a proxy for lifestyle disease surveillance. Existing studies, however, are prone to at least one of the following issues: ad-hoc keyword selection, overfitting, insufficient predictive evaluation, lack of generalization, and failure to compare against trivial baselines.

Objective: The aims of this study were to (1) employ a corrective approach improving previous methods; (2) study the key limitations in using Google Trends for lifestyle disease surveillance; and (3) test the generalizability of our methodology to other countries beyond the United States.

Methods: For each of the target variables (diabetes, obesity, and exercise), prevalence rates were collected. After a rigorous keyword selection process, data from Google Trends were collected. These data were denormalized to form spatio-temporal indices. L1-regularized regression models were trained to predict prevalence rates from denormalized Google Trends indices. Models were tested on a held-out set and compared against baselines from the literature as well as a trivial last year equals this year baseline. A similar analysis was done using a multivariate spatio-temporal model where the previous year's prevalence was included as a covariate. This model was modified to create a time-lagged regression analysis framework. Finally, a hierarchical time-lagged multivariate spatio-temporal model was created to account for subnational trends in the data. The model trained on US data was, then, applied in a transfer learning framework to Canada.

Results: In the US context, our proposed models beat the performances of the prior work, as well as the trivial baselines. In terms of the mean absolute error (MAE), the best of our proposed models yields 24% improvement (0.72-0.55; $P<.001$) for diabetes; 18% improvement (1.20-0.99; $P=.001$) for obesity, and 34% improvement (2.89-1.95; $P<.001$) for exercise. Our proposed across-country transfer learning framework also shows promising results with an average Spearman and Pearson correlation of 0.70 for diabetes and 0.90 and 0.91 for obesity, respectively.

Conclusions: Although our proposed models beat the baselines, we find the modeling of lifestyle diseases to be a challenging problem, one that requires an abundance of data as well as creative modeling strategies. In doing so, this study shows a low-to-moderate validity of Google Trends in the context of lifestyle disease surveillance, even when applying novel corrective approaches, including a proposed denormalization scheme. We envision qualitative analyses to be a more practical use of Google Trends in the context of lifestyle disease surveillance. For the quantitative analyses, the highest utility of using Google Trends is in the context of transfer learning where low-resource countries could benefit from high-resource countries by using proxy models.

(J Med Internet Res 2020;22(1):e13347) doi:10.2196/13347

KEYWORDS

noncommunicable diseases; lifestyle disease surveillance; infodemiology; infoveillance; Google Trends; Web search; nowcasting; public health; digital epidemiology

Introduction

Background and Prior Work

Public health surveillance is the systematic collection, analysis, and interpretation of health-related data to be used by those responsible for preventing and controlling disease and injury [1]. One of the most common examples of public health surveillance involves what is known as *disease surveillance*. Disease surveillance is traditionally accomplished through a system of manual surveys, or mandatory reporting by the doctors to the government. However, such a system is costly, prone to missing new and rare events, and has a high time lag. Hence, in the past decade, there has been an increase in the use of Web-based data for disease surveillance with the goal of supplementing, not replacing, traditional methods.

One of the first applications of Web-based disease surveillance was the tracking of influenza using Web-based search behavior. A seminal study on this was carried out by Ginsberg et al [2] in the creation of Google Flu Trends (GFT). The purpose of GFT was to monitor health-seeking behavior by analyzing Google search queries to track influenza-like illness (ILI) in a population. Although, the system shut down in 2015 for overestimating the influenza epidemics [3], it started a whole line of research using Google Trends to nowcast anything from US presidential elections [4], to fertility rates [5], stock prices [6], box office sales [7], and other economic indicators [8]. But more than anything, it set a precedent for using Web search data for nowcasting and monitoring diseases [9].

Most of the work on disease surveillance focuses on *fast-moving* infectious or communicable diseases where the goal is to predict the disease outbreak as early as possible [10]. In this domain, previous work includes using data for predicting influenza in South China [11], using Google Trends to determine relationship between sexually transmitted infection (STI)–related search engine trends and STI rates [12], and the use of other search engines such as Yahoo for surveillance of influenza in the United States [13].

However, noncommunicable diseases (NCDs) or lifestyle diseases, such as obesity, smoking, diabetes, depression, or lack of physical activity, account for a far larger share of both the US and global health care system's cost. According to the World Health Organization, lifestyle diseases are responsible for around 70% of all deaths globally every year [14]. In the United States, the economic burden of obesity-related diseases alone is around US $190 billion [15]. This makes tracking of these diseases important for timely allocation of resources and implementation of interventions such as taxation, policy changes, or public health campaigns. However, the current traditional methods of surveillance used in the United States are costly, labor-intensive, and time-consuming. In the international setting, most countries do not even have such surveillance systems, and lack data and statistics about population behavior and disease risk factors. This has triggered researchers to use data on the Web, and specifically Web search activity as a proxy to predict the prevalence of NCDs.

Google Trends is one of the most popular tools for analyzing Web search activity. Research in the domain of using Google Trends covers anything from the prediction of suicide risks, depression, shared migraine experiences, and stress in a population [16-21] to monitoring of nonsuicidal self-injury rates [22], correlational studies between Google Trends and actual suicide risk and rates [23-27], influence of seasons on the incidences of depression [28], study of psychological and social factors affecting internet searches related to suicide [29], seasonality in seeking of mental health information [30], low validity of Google Trends in forecasting suicidal risk [31], the infoveillance of cancer incidence rates [32,33], mortality rates [34], obesity [35], diabetes [36], dental caries [37-39], the behavioral forecasting and awareness of alcohol consumption rates, and drugs [17], seasonal variation in ophthalmology-related diseases [40], geographical variance in stroke prevalence [41], and a multitude of other relevant literature, a survey of which can be found in [9].

Existing Challenges in Modeling Noncommunicable Diseases With Google Trends Data

As illustrated above, the literature on the infodemiology and infoveillance of NCDs is extensive and diverse. However, almost all of it suffers from a fundamental problem when trying to build models—the scarcity of data. This is primarily a consequence of the *slow-moving* nature of most of the NCDs, as well as the lack of resources to conduct finer temporal surveillance. As a result, the surveillance data for lifestyle diseases are typically available on an annual basis as opposed to weekly basis for ILI. It is important to note that even if the finer temporal surveillance was possible, owing to the *slow-moving* nature of the NCDs, there would not be discernible changes in the data points to conduct any useful analysis as, for example, obesity rates are unlikely to change week over week. Notwithstanding the data scarcity issues in the temporal surveillance, most of the background literature still falls back on using time-series data to do some version of correlational studies. These analyses span across countries, states, cities, or metro areas. However, because Google Trends data are only available starting 2004, *most* of the temporal correlational studies for NCDs will have only as many data points as years passed since 2004: a maximum of 15 data points for a given location. This results in nonrobust time-series-based approaches as they are subject to overfitting. Another major limitation of these studies is the ad-hoc keyword selection. More concretely, most of the aforementioned literature resorts to using hand-picked keywords as features for the target variable at hand. This introduces an a priori bias and subjectivity in the predictive system, as well as the evaluation.

To deal with the first problem on the limitations of the temporal data, a few studies have tried to use state-level data to fit US national-level trends. In this regard, a relevant research was

carried out by Sarigul and Rui [35] on nowcasting obesity. They used manually selected terms and their correlations to predict regional obesity prevalence by modeling regional as well as temporal variations in the Google Trends data. However, their methods use ad-hoc keyword selection and tend to use a fitting approach without predictive out-of-sample evaluations.

Another relevant study has used spatial data to predict temporal trends. This research was carried out by Nguyen et al [42] and uses linear regression model along with Lasso regularization, modeling regional variation for different NCDs to predict prevalence by state for a particular target year. One novel contribution of their study is the semiautomated keyword selection process using semantically related terms as keywords. Unlike other studies, they perform an out-of-sample evaluation. However, one of their key shortcomings is the lack of an appropriate denormalization model as explained as follows: Although they train a model across space, they, then, apply it across time, without accounting for the fact that, within a given year, the spatial data are independently normalized by Google. Hence, although the individual spatial trends might be appropriate to track with these data, the national-level model would miss out on national trends in time. As an example, if the national search intensity for terms predictive of NCDs was to double from 2014 to 2015 with the relative spatial distribution remaining constant, then spatial data alone would not pick up such a temporal trend. In fact, it would treat 2014 exactly as 2015. To account for such national temporal variations, Phillips et al's study on the relationship between state-level search behavior and the cancer incidence in the United States [32] uses time in years as a continuous covariate to control for temporal trends. However, due to the fundamental intricacies of the way Google normalizes its data, discussed in the next section, such an approach is not likely to generalize.

A very small subset of the background literature [30,31,37-40,43] explicitly tests the application of their models to other countries or geographic regions. In essence, most of the previously proposed surveillance models or techniques remain inconclusive in their generalizability to other spatial reference frames. To correct for this, we explore the utility of our approach in the international setting by using a transfer learning framework.

Finally, *all* of the literature on the surveillance of NCDs using Google Trends leaves out obvious yet important evaluation criteria: a comparison of their results to the trivial *last year equals this year* baseline. This is going to be one of the key themes of this paper while we evaluate the validity of Google Trends to predict the prevalence of NCDs.

A brief summary of the comparison of previous literature on the lifestyle disease surveillance using Google Trends in relation to our contribution across several metrics is presented in Table 1. The first column represents the literature surveyed; the second column represents if the literature used any sort of automation in terms of keyword selection; the third column surveys the type of data used in the study; the fourth column represents the inclusion of data denormalization if applicable; the fifth column surveys the type of evaluation used; the sixth column shows if the study compared its evaluation with a trivial baseline; the seventh column describes the geographical setting the study was based on; and the eighth column describes if any secondary evaluation of the proposed methodology is shown for a different geographical setting. In terms of the cell values, *x* represents missing; *N/A* represents not applicable; and ✓ represents available.

Table 1. A survey and comparison of previous literature across different metrics.

Studies	Bootstrapping keyword selection	GT data type	Data denormalization	Predictive evaluation	Comparison to trivial baseline	Geographical setting	Generalizability to other geographical setting
Leffler et al [40]	x[a]	Temporal	N/A[b]	In-sample	N/A	United States, the United Kingdom, Canada, and Australia	✓[c]
Yang et al [28]	x	Temporal	N/A	In-sample	N/A	Worldwide	x
McCarthy [18]	x	Temporal	N/A	In-sample	x	United States	x
Hagihara et al [24]	x	Temporal	N/A	In-sample	N/A	Japan	x
Sueki [25]	x	Temporal	N/A	In-sample	N/A	Japan	x
Walcott et al [41]	x	State-level	N/A	In-sample	N/A	United States	x
Yang et al [23]	x	Temporal	N/A	In-sample	N/A	Taipei City, Taiwan	x
Ayers [30]	x	Temporal	N/A	In-sample	N/A	United States, and Australia	✓
Bragazzi [22]	x	Temporal	N/A	In-sample	N/A	Italy	x
Braun and Harréus [44]	x	Temporal	N/A	In-sample	N/A	Germany	x
Breyer and Eisenberg [36]	x	Temporal	N/A	In-sample	N/A	United States	x
Gunn III and Lester [21]	x	State-level	N/A	In-sample	x	United States	x
Ingram Plante [43]	x	Temporal	N/A	In-sample	N/A	United States, Australia, Germany, the United Kingdom, and Canada	✓
Willard and Nguyen [45]	x	State-level	N/A	In-sample	N/A	United States	x
Bragazzi [46]	x	Temporal	N/A	In-sample	x	Italy	x
Brigo et al [47]	x	Temporal	N/A	In-sample	N/A	Worldwide	x
Bruckner et al [26]	x	Temporal	N/A	In-sample	N/A	England and Wales	x
Sarigul et al [35]	x	State-level	x	In-sample	x	United States	x
Song et al [29]	x	Temporal	N/A	In-sample	N/A	Korea	x
Nguyen et al [42]	✓	State-level	x	Out-of-sample	x	United States	x
Wang et al [48]	✓	Temporal	N/A	In-sample	x	Taiwan	x
Ma-Kellams et al [19]	x	State-level	N/A	In-sample	x	United States	x
Parker et al [17]	x	State-level	x	Out-of-sample	x	United States	x

Studies	Bootstrapping keyword selection	GT data type	Data denormalization	Predictive evaluation	Comparison to trivial baseline	Geographical setting	Generalizability to other geographical setting
Burns et al [16]	x	Temporal	N/A	In-sample	N/A	United States	x
Cervellin et al [49]	x	Temporal	N/A	In-sample	x	Italy	x
Hassid et al [50]	x	Temporal	N/A	In-sample	N/A	United States	x
Lotto et al [38]	✓	Temporal	x	In-sample	N/A	United States, United Kingdom, Australia, and Brazil	✓
Ojala et al [5]	✓	State-level, temporal	N/A	Out-of-sample	x	United States	x
Ricketts and Silva [34]	x	Temporal	N/A	In-sample	x	United States	x
Tran et al [31]	✓	Temporal	N/A	In-sample	N/A	United States, Germany, Austria, and Switzerland	✓
Wehner et al [33]	x	Temporal	N/A	In-sample	x	United States	x
Aguirre et al [37]	✓	Temporal	x	In-sample	N/A	United States, United Kingdom, Germany, Brazil, France, India, Italy, Japan	✓
Arendt [27]	x	Temporal	N/A	In-sample	x	Worldwide	x
Chandler [20]	x	State-level	x	In-sample	x	United States	x
Coogan et al [51]	x	Temporal	N/A	In-sample	x	Australia	x
Phillips et al [32]	x	State-level	x	In-sample	x	United States	x
Cruvinel et al [39]	✓	Temporal	x	In-sample	N/A	10 South American Countries	✓
This study [d]	✓	*State-level, temporal*	✓	*Out-of-sample*	✓	*United States*	✓

[a]x: missing.

[b]Not applicable.

[c]✓: available.

[d]The values in italics signify how our study compares to those from the past literature across different metrics.

In summary, we identified the following key issues in the aforementioned background literature:

1. Ad-hoc keyword selection.
2. Overfitted temporal analysis.
3. Spatial analysis without appropriate denormalization.
4. Insufficient predictive evaluation.
5. Lack of evidence for generalization to other countries.
6. Failure to compare results to trivial baselines.

Study Objectives

The insufficient evaluation metrics, and methodological errors in the background review, set out a motivation to validate the use of Google Trends for nowcasting NCDs. Hence, in this study, we had 3 key objectives:

1. To use a corrective approach to first rectify the methodological shortcomings of the previous literature.
2. To study the limitations and promises of Google Trends in the context of its accuracy and robustness to predict national lifestyle disease trends.
3. To experimentally test the generalizability of this approach to other similar countries.

Methods

Study Design

The methods we used for our study differ from previous work, as explained above, in (1) how we select search terms for Google Trends to limit cherry picking, (2) how we denormalize Google Trends data to overcome certain limitations, (3) the focus on an out-of-sample evaluation rather than in-sample model fit, (4) the inclusion of a trivial baseline for comparison, and (5) a transfer learning setup to evaluate cross-country generalizability.

Terminology

To facilitate understanding of the description of our methodology, we have defined a set of key terms that we used frequently in this paper. These terms and their definitions are as follows:

Offline Target Variable

This refers to the variable of interest that we are monitoring. In our case, we monitored diabetes, obesity, and exercise.

Offline Data

For any offline target variable (such as diabetes), offline data are the actual regional or national prevalence of the condition. Although *offline target variable* refers to the name of the variable we are monitoring—say, *Diabetes*—*offline data* refers to actual numerical values pertaining to that variable—say, US state-level diabetes prevalence rates in 2014.

Spatial Data

For the purpose of this paper, spatial data refer to Google Trends' Web search intensity for a given year and a particular keyword normalized across different US states.

Temporal Data

Temporal data refer to Google Trends' US Web search intensity for a particular keyword normalized across different years.

Offline Data Collection

For each of the target variables, we collected the offline data across 15 years from 2004 to 2018 each year separately. This includes data for the 50 states (including Washington, DC and excluding Hawaii as offline data for Hawaii were unavailable for the year of 2004). For prevalence rates, we used the Center for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS) [52].

Keyword Selection

In this phase, we used 3 tools for the keyword selection: (1) Google Correlate, (2) related search queries, and (3) Semantic Link, a Web-based service to find related terms. A subset of the resulting keywords from each of the 3 sources is presented in Table 2.

We bootstrapped the keyword selection process as follows, starting from a set of seed terms.

Table 2. The subset of unpruned keywords for different target variables.

Target variable	Google Correlate	Semantic Link	Related Queries
Diabetes	when i get up	insulin	diabetes symptoms
	sell avon	polyphagia	signs of diabetes
	medicine for dogs	ketoacidosis	prediabetes
	very weak	cholesterol	icd 10
	sugar level	hypertension	icd 10 type 2 diabetes
Obesity	catherines.com	abdominal	food delivery near me
	dresses plus size	anorexia	lose fat
	sims 3 games	BMI	myfitnesspal
	lose 100 pounds	appetite	indeed.com
	dresses plus	ADHD	pizza delivery
Exercise	transportation options	exercises	my fitness pal
	best bike	aerobic	workout
	bike laws	jogging	iPod
	bike repair	gyms	quinoa gluten free
	bike frame size	muscles	how to exercise

Seed Terms

For each target variable, we used 1 or 2 seed terms. For diabetes, we chose *diabetes* and *diabetic*, for obesity, we chose *obesity* and *obese*, and for exercise, we chose *exercise*. These seed terms were, then, used to generate other cooccurring terms in English Wikipedia using Semantic Link [53] as mentioned in the study by Nguyen et al [42]. Other methods to enhance related terms, such as those described in the study by Lampos et al [54] on using word embeddings could also be used.

Google Correlate

Google Correlate is a tool that takes either a temporal or a spatial series as input and returns a ranked list of Web search queries

that are correlated across time or space [55]. For our purposes, we used the offline data for the year 2015 to determine top 30 to 40 keywords that strongly correlate across all the US states including Washington, D.C. We, then, pruned many of those keywords as has been described later.

Related Search Queries

We used a combination of the keywords selected in the aforementioned methods in Google Trends to output-related search queries.

Pruning

One of the most common methods to avoid overfitting is by using dimensionality reduction or by using regularized models. However, these methods do not guard against keywords with spurious correlations. For example, the query *sims 3 games* in Table 2 is one of the top 10 spatially correlated keywords for obesity. A less obvious example is the search term, *catherines.com*. Catherines is a store selling plus-size clothing and so search volume for *catherines.com* has an arguable causal connection to obesity. However, its search volume is also tied to its market share, which can change over time. As such, it might be a robust feature for a spatial-only model, but we decided to remove such *branded* search terms for temporal analysis. Finally, apart from removing nonsensical and branded search terms, we also removed any keywords with low search volume. All these selections are made in an effort to reduce overfitting and increase model robustness across time. A list of all the postpruning keywords used for this study for each of the target variables can be found in Multimedia Appendix 1.

Google Trends Data Collection

We used Google Trends to collect 2 kinds of data: spatial and temporal. For spatial data, we used each keyword-year combination as a query to Google Trends to get across-state data, that is, 50 data points for each keyword and each of the 15 years, 2004-2018. For temporal data collection, we collected data across the 15 years at the US national level rather than at the state level. This was done to reduce the required data collection effort by a factor of 50. Note that the state-level temporal trends are implicitly collected already as we have (1) state-level relative volumes within a given year, as well as (2) US national-level temporal trends across the years. We have explained how we combine these 2 types of data points to create a spatio-temporal model in following sections.

As Google Trends does not provide an official application programming interface, other than their *export as .csv* option, we made use of Python's *pytrends* [56] package, which can be used to retrieve data from Google Trends. Regardless of which method is used to obtain the data, one caveat is that Google Trends' data are not stable and that repeatedly asking for the same data can return different results. Concretely, since Google search volume index is calculated by a sampling method, the results even for historic data can fluctuate day to day [31,57]. To limit such fluctuations and instability in the search data, similar to [31], we sample and average each data point 10 times across time with a gap of a day between each sample for the United States, and 3 times across each sample for Canada.

Another important detail is that Google Trends results for an individual term such as *diabetes* include search phrases such as *diabetes insulin* or *insulin diabetes* [58]. This can create collinearity for the results for different terms, which has to be taken into account when and if doing a post hoc feature analysis.

Google Trends Data Normalization

For both privacy and business reasons, Google Trends does not show absolute search volumes but only *normalized* search intensity, and it is important to understand the process of this normalization [59]. One of our contributions is a data calibration mechanism, which is based on a proper understanding of the underlying normalization procedure.

Concretely, search intensity is different from absolute search volume in that it measures the *relative* interest in a search term, that is, the fraction of all searches in the reference temporal or spatial unit. One desired consequence of this is that as Google's user base grows over time, the search intensity does not trivially increase, making it potentially comparable across time.

Another point to understand is that there is a certain interplay between the search terms. For example, if the search volume for the keyword *justin bieber* was to go up 10-fold, with everything else remaining constant, then the *relative* search intensity for other terms would still drop (slightly), even though their *absolute* search volume remains unchanged.

The search intensity is normalized across time or across space depending on the mode of data collection. For example, when collecting temporal data across several years for the query *tomacco*, Google returns data such that the period with the highest relative search intensity corresponds to an arbitrary reference value of 100. All other temporal units are normalized with respect to this absolute maximum of 100, meaning that a value of 30 means that in a corresponding time unit, the relative fraction of searches matching the criteria was only 30% of what it is during the peak. Similarly, when getting search data across spatial units such as US states, the state with the highest search intensity is assigned a value of 100, and all other spatial units are normalized relative to the search intensity of that peak location.

In our setting, we combined data across both space and time. The primary reason behind doing that is to increase the number of training instances in an effort to compensate for the slow-moving nature of the NCDs. This, in turn, helps us to learn a national generalizable model. Combining spatial and temporal data requires undoing Google's normalization for the following reason: within each year, data are normalized independently across space. Thus, the value of 100 in year 2014 for the state of California cannot be compared with the value of 100 in the year 2015 for the state of Texas. From 2014 to 2015, the overall search volume may have gone up or down, and spatial data alone do not reveal such information. By appropriately combining the spatial data with temporal trends, we are able to effectively undo Google's normalization to correctly juxtapose the data for 2014 next to the data for 2015 such that a relative increase in numbers actually corresponds to a relative increase in search intensity.

To reconstruct the state-level contribution to the national value, we need to take into account the actual absolute search volume from each of the states. As those data are not available to us, we approximated that using the following 2 steps:

(1) To undo the effect of spatial and temporal normalizations, we chose 2004 as a reference year *r* to rescale and denormalize year-state values as in equation (1) in Figure 1.

Here is the denormalized scalar value for year *y* and state *s*. Furthermore, *G* signifies that the data were collected via Google Trends, G_l represents that data are spatial (where *l* stands for location), that is, normalized across the US states by Google, represents the year, *s* represents the state, and $G_l(x_{ys})$ represents a single state-level data point for the year *y* obtained from Google Trends. G_t represents that data are temporal (where *t* stands for time), $G_t(z_y)$ represents the value of the corresponding keyword at the national level in year *y* across time, that is, normalized across the different years, $G_t(z_r)$ represents the across-time value of the corresponding keyword in the reference year *r*, where for our purposes, *r=2004*. $\sum^n_i G_l(x_{ri})$ represents the sum of the regional distribution of the corresponding keyword in the reference year 2004 where *n* is the number of states. $\sum^n_i G_l(x_{yi})$ represents the sum of the regional distribution of the corresponding keyword in the year y.

(2) Another important insight to realize is that different regions in the United States contribute differently to the US national trends because of differences in absolute search volume. Regions with large populations, and large numbers of Google search users, such as California or New York, will have more of an impact on the national trend than regions with small populations. However, the relative search intensities normalize for different numbers of issued Google searches. Hence, to debias our data on the population level, after following step 1, we adjusted each value by a product of the population in each state multiplied by the internet penetration to get an approximate number of Google search users in each state. We collected the internet penetration rates from the BRFSS site [52]. Note that we do not need to have an absolute number of Google users. For our method to work correctly, all that matters is that we have a relative multiple of that (unknown) number.

As both state-level populations and internet penetration can change over time, ideally, we would want to use different correction factors for each year. However, as we observed that both population sizes and internet penetration rates increased fairly uniformly across all states, the correction factors for different years were correlated at the level of *approximately .99*. For this reason, we chose to apply only a single, static, set of state-level correction factors from the year 2015.

After following the denormalization procedure, we were left with a matrix where each year-state value can be compared with each other year-state value in a meaningful manner. This allowed us to, effectively, multiply our training data across different years and different states.

With steps 1 and 2, our final formula was as shown in equation (2) in Figure 1, where all the terms are same as the equation (1). P_{ri} represents the population size for the reference year *r* and state *i*. P_{yi} represents the population size for the year *y* and state *i*. I_{ri} represents the internet penetration for the reference year *r* and state *i*. I_{yi} represents the internet penetration for the year *y* and state *i*

Note that the final output $\hat{x}_{ys}$ is a scalar, indexed by both year *y* and US state *s*.

For an example-based explanation of the denormalization process, see Multimedia Appendix 2 [60].

Figure 1. The equations for the proposed denormalization framework.

Number	Equation
(1)	$\hat{x}_{ys} = G_l(x_{ys}) * \frac{G_t(z_y)}{G_t(z_r)} * \frac{\sum_i^n G_l(x_{ri})}{\sum_i^n G_l(x_{yi})}$
(2)	$\hat{x}_{ys} = G_l(x_{ys}) * \frac{G_t(z_y)}{G_t(z_r)} * \frac{\sum_i^n G_l(x_{ri}) * P_{ri} * I_{ri}}{\sum_i^n G_l(x_{yi}) * P_{yi} * I_{yi}}$

Regression Modeling

Even after our approach for obtaining our year-state data matrix, we did not have sufficient data to fit complex prediction models such as deep neural networks [61]. Hence, we fit (regularized) linear regression models to predict slow-moving trends such as diabetes, obesity, and exercise rates for the 50 states in the United States. We used Python's scikit [62] library to fit our linear regression models. Concretely, given y as the ground truth and $\hat{y}$ as the prediction, we fit a model of the form:

$$\hat{y}=wx+b \quad (1)$$

by minimizing the loss L

$$L=(y-\hat{y})^2+\lambda \quad (2)$$

where x represents the feature set, w represents the unknown parameters, and b represents the bias.

To avoid overfitting, and to yield a simpler and interpretable model, we used Lasso [63] as the shrinkage, which combines ordinary least-square regression with an L-1 regularization. We also experimented with Ridge regression (without consulting the testing set) [64] but the cross-validated performance on the training set was similar, and Lasso regression gave a sparser model. Lasso requires a regularization parameter λ to govern the trade-off between more complex models with better performance on the test data (=small λ) and simpler models with potentially better performance on unseen data (=larger λ).

In this paper, the optimal λ was determined by using k-fold cross validation, optimizing for the negative mean squared error. Feature values were standardized to have zero mean and unit variance. For our purposes, we used *k=12* corresponding to the number of years in the training set (2005-2016).

Training, Validation, and Testing Phase

For the training phase, we used data from 2005 to 2016. We trained and cross validated our model using a k-fold cross-validation where *k=12*. For diabetes, obesity, and exercise, each year had 50 data points, one for each state (including the District of Columbia and excluding the state of Hawaii). In total, the training was performed using 600 data points. Note that we did not include data points from 2004 in our experiments. This is done to have a consistent training and test set as our proposed methods require a *lag* where data from 2004 is used in creating a feature vector for 2005, and similarly for following years.

We later tested each of our models trained on data from 2005 to 2016 on the years 2017 and 2018, and calculated the mean absolute error (MAE), root mean squared error (RMSE), symmetric mean absolute percentage error (SMAPE) [65], Spearman's Correlation Coefficient (rho), and Pearson's Correlation Coefficient (R).

We emphasize that, during development, we never looked at the results on the test set, and so none of our design decisions were influenced by them to safeguard against implicit overfitting. This is one of the key shortcomings of the previous literature, most of which employ in-sample evaluation.

Alternate Approaches

We performed 6 different approaches for each of the 3 target variables for the US region. We have defined these 6 experiments in detail as follows.

Trivial Baseline

For evaluation purposes, we used the trivial *last year equals this year* baseline. With this in mind, we used 2016's prevalence as a prediction for 2017, and 2017's prevalence as a prediction for 2018. We, then, computed the MAE, RMSE, SMAPE, rho, and R, to be used as baseline for evaluation purposes.

Spatial Model

We extended the methodology presented in Nguyen et al's paper [42] of applying a national spatial-only model to temporal dimension. This becomes our secondary baseline.

Spatio-Temporal Model

This experiment is based on our main methodological contribution where we used a corrective approach to first neutralize the effect of Google's normalization as a preprocessing step, and, then, trained the model. We called it *Spatio-Temporal* to signify the use of both spatial and temporal Google Trends data.

Multivariate Model

To boost the performance of our spatio-temporal model, we used the *trivial baseline* as a covariate. More concretely, we extended our spatio-temporal model to use the actual prevalence of the previous year as an auxiliary feature to predict the prevalence of the current year. We called the extended model *multivariate* to signify the inclusion of the covariate.

Lagged Multivariate Model

While training the previous model, we made a crude assumption that the population search behavior for any particular year is correlated to the prevalence of that year. However, we realized that it may be possible for the search behavior of any year to be predictive of the next year. As an example, search behavior in 2017 may be more predictive of 2018 prevalence than of 2017. To test this theory, we experimented by shifting our time window for the multivariate regression.

Hierarchical Lagged Multivariate Model

One of the simplifying assumptions we made while training the previous models is that the predictive pattern of Google Trends is the same across all states. This assumption may, however, not be valid. In particular, each state might have a different base prevalence rate for the health condition being modeled. To incorporate subnational bias terms, we explicitly included the state ID as a covariate. We did this by extending the feature vector to include a one-hot vector of 50 states. By doing this, we implicitly modeled a hierarchical distribution where we considered national trends, as well as subnational trends.

Transfer Learning

We tested the generalizability of our methodology across countries by conducting 2 further experiments for the prevalence of diabetes and obesity in Canada. We collected the Google Trends data for Canada in a similar fashion as we did for the

United States. The offline health statistics for diabetes and obesity were collected from the Statistics Canada site [66]. One limitation pertaining to these statistics was that, starting from 2015, the collection strategy and the design of the sampling process for synthesizing statistics has changed, rendering pre-2015 data incomparable to post-2015. Therefore, we dealt with these 2 periods of data separately in our experiments. A second point to note is that diabetes and obesity statistics for Canada are not available for 2004 and 2006. We, therefore, collected the offline data from 2007 onward only. Owing to these 2 limitations and the need to have 1 separate year for the lagged models, our training set for Canada included data from 2008 to 2012 with each year containing 10 data points, 1 per Canadian province. We used data from 2013 and 2014 as a test set. Owing to the change in sampling methods and the limited number of years, we did not train a separate model for 2016-2018. However, we reported the results for these years as a test set.

Cross-Country Generalizability of the Method

In the first set of experiments in the context of transfer learning, we used the same set of experiments that we conducted for the United States, to make predictions for Canada.

Cross-Country Generalizability of the Model

For the second set of experiments, we trained our lagged multivariate model on the United States on 14 years (from 2005 to 2018) to test across Canada. The purpose for this experiment was 2-fold. First, we wanted to test the generalizability of our trained model across other *similar* countries by using Canada as a proxy. Second, we wanted to test the reliability of applying the models in such a fashion to other countries where offline data might not be available. We have briefly described the importance of this step later in the Discussion section.

Results

Evaluation

For the evaluation of our experiments, we measured the performance of different models over 5 metrics: MAE, RMSE, SMAPE [65], rho, and R. To compute the correlation coefficients, we concatenated the predictions across all the test years and used those to compute a global correlation.

To test for statistically significant improvements in the MAE, we conducted a one-sided paired *t* test across each set of experiments in relation to the trivial baseline. We also computed statistical tests for each extension such that the spatio-temporal model gets compared with the spatial model, the multivariate spatio-temporal model gets compared with the spatio-temporal model, and so on. We do this to evaluate the gain, if any, obtained by each extension.

US Based Models

A detailed evaluation of our experiments on the United States can be found in Table 3. The best way to interpret the results is to read the values for each statistic from left to right.

Table 3. A detailed evaluation of 5 different experiments across the 3 target variables for the region of the United States.

Target variable	Trivial baseline	Spatial model	Spatio-temporal model	Multivariate spatio-temporal model	Lagged multivariate spatio-temporal model	Hierarchical lagged multivariate spatio-temporal model
Diabetes						
MAE[a]	0.72	0.81	*0.72* [b]	*0.65* [c]	*0.62* [c]	*0.55* [c]
RMSE[d]	0.92	1.0	*0.91*	*0.81* [c]	*0.80* [c]	*0.72* [c]
SMAPE[e]	6.94	7.63	*7.13*	*6.23* [c]	*6.04* [c]	*5.24* [c]
Spearman rho	0.87	*0.89* [c]	0.87	*0.88* [c]	*0.89* [c]	*0.93* [c]
Pearson R	0.90	0.90	0.88	*0.91* [c]	0.91[c]	*0.94* [c]
Obesity						
MAE	1.20	2.81	*2.09*	*1.24*	*0.99* [c]	1.08[c]
RMSE	1.55	3.28	*2.56*	*1.59*	*1.33* [c]	1.40[c]
SMAPE	3.88	9.31	*6.96*	*4.03*	*3.22* [c]	3.51[c]
Spearman rho	0.93	0.87	0.85	*0.94* [c]	0.93	*0.95* [c]
Pearson	0.93	0.86	0.86	*0.94* [c]	0.94[c]	*0.95* [c]
Exercise						
MAE	2.89	*2.32* [c]	3.12	*2.47* [c]	*2.36* [c]	*1.95* [c]
RMSE	3.32	*2.75* [c]	3.75	*2.90* [c]	*2.83* [c]	*2.40* [c]
SMAPE	3.85	*3.11* [c]	4.11	*3.32* [c]	*3.16* [c]	*2.62* [c]
Spearman rho	0.68	*0.73* [c]	*0.81* [c]	0.71[c]	*0.77* [c]	*0.80* [c]
Pearson R	0.69	*0.74* [c]	*0.80* [c]	0.72[c]	*0.78* [c]	*0.81* [c]

[a]MAE: mean absolute error.

bThe values in italics signify an improvement in the performance in comparison to the previous method.

[c]The method beat the trivial baseline.

[d]RMSE: root mean squared error.

[e]SMAPE: symmetric mean absolute percentage error.

Toward our first research objective to correct the methodological shortcomings of the previous literature, we developed a *spatio-temporal model* for nowcasting lifestyle diseases. We observed that the spatio-temporal model performs better than the spatial model for diabetes and obesity but not for exercise. The performance improvement over the spatial method for diabetes in terms of the MAE (0.81-0.72) and RMSE (0.91-0.81) was 11% (P=.06), whereas the improvement for obesity was 26% in MAE (2.81-2.09) and 22% in RMSE (3.28-2.56; P<.001). However, only the results for obesity are significant. Both of these improvements in MAE are statistically significant.

To improve upon the spatio-temporal model, we, then, trained a *multivariate spatio-temporal model*, by using the previous year's prevalence as a covariate in combination with features from Google Trends. This improves the performance over the spatio-temporal model decreasing the error for diabetes by 10% in MAE (0.72-0.65) and 12% in RMSE (0.91-0.81), obesity by 41% in MAE (2.09-1.24) and 38% in RMSE (2.59-1.59), and exercise by 21% in MAE (3.12-2.47) and 23% in RMSE (3.75-2.90). Of these improvements, diabetes was not significant (P=.08), but both obesity and exercise were statistically significant (P<.001 and P=.001, respectively).

In the next set of experiments, we shifted our dependent response variable by 1 year to test for a time lag in the impact of search behavior on disease statistics. We called this the *lagged multivariate spatio-temporal model*. In comparison with the multivariate spatio-temporal model, this model improves performance on diabetes by 5% in MAE (0.65-0.62) and 1% in RMSE (0.81-0.80), on obesity by 20% in MAE (1.24-0.99) and 16% in RMSE (1.59-1.33), and on exercise by 4% in MAE (2.47-2.36) and 2% in RMSE (2.90-2.83). The improvement was statistically significant for diabetes and obesity (P=.02 and P<.001), but not for exercise (P=.17).

Finally, to account for subnational trends, we trained a *hierarchical lagged multivariate spatio-temporal model* where we included the state ID as a covariate. In comparison with the lagged model, we got a performance improvement on diabetes and exercise, but a decrease in performance on obesity. For diabetes, the MAE improves by 11% (0.62-0.55) and RMSE by 10% (0.80-0.72). For exercise, the MAE improves by 17%

(2.36-1.95) and RMSE by 15% (2.83-2.40). Whereas for obesity, the performance deteriorates by 8% in MAE (0.99-1.08) and by 5% in RMSE (1.33-1.40). For both diabetes and exercise, the performance improvement was statistically significant (P=.01 and P=.02).

In terms of the overall improvement over the trivial baseline, our best models (ie, hierarchical lagged multivariate spatio-temporal model for diabetes, and exercise, and lagged multivariate spatio-temporal model for obesity) result in 24% improvement in MAE (0.72-0.55) and a 22% improvement in RMSE (0.92-0.72) for diabetes; 18% improvement in MAE (1.20-0.99) and a 14% improvement in RMSE (1.55-1.33) for obesity, and 34% improvement in MAE (2.89-1.95) and a 28% improvement in RMSE (3.3-2.40) for exercise. All of these improvements are found to be statistically significant (P<.001 for diabetes, P=.001 for obesity, and P<.001 for exercise). The SMAPE, Spearman rho, and Pearson R follow similar trends.

Transfer Learning

For the experiments for our transfer learning framework, we evaluated (1) cross-country generalizability of the method (training and evaluating models just for Canada) and (2) cross-country generalizability of the model (taking a trained-on-US model and evaluating it for Canada). We summarized the results achieved for each of the proposed methods in Tables 4-6.

Whereas for the US case, most of our methods beat the trivial baseline, this is not the case for either of the 2 transfer settings. For the approach of retraining a Canada-specific model, we hypothesized that this is due to data scarcity with fewer data points for a given year (10 vs 50) and fewer years to train on (2008-2012 vs 2005-2016) being available compared with the US case. For some of the experiments (eg, diabetes model trained on 2008-2014 and tested on 2016-2018) where we beat the baseline, the improvements were only marginal. A detailed set of evaluation can be found in Tables 4 and 5.

Although the approach of applying the trained-on-US model to Canada fails to beat the trivial this-year-is-same-as-last-year baseline, we expect the highest utility of such approaches to be reliable when health statistics for the target country are unavailable. For this scenario, our results showed promise by achieving an average Spearman and Pearson correlation of 0.70 for diabetes, and 0.91 and 0.92 for obesity by using a pretrained model of a *similar* country (which, in our case, is the United States). We believe that these results are encouraging and worth replicating for developing countries.

Table 6 shows the results for the transfer learning framework across the 2 target variables for Canada for generalizability of the US trained model. We report the performance of the US-based model trained for the years 2005 to 2018 on Canada for the test years 2008-2014. We separately also report the performance of the model on the years 2016-2018 and 2012-2013 for the readers to compare the performance of the model trained on the United States to the model trained on Canada, both tested on the same test set.

Table 4. Results for the transfer learning framework across the 2 target variables for Canada for generalizability of the method trained over the years 2008-2012, and tested on the years 2013 and 2014.

Target variable	Trivial baseline	Spatial	Spatio-temporal model	Multivariate spatio-temporal model	Lagged multivariate spatio-temporal model	Hierarchical lagged multivariate spatio-temporal model
Diabetes						
MAE[a]	0.54	0.80	0.83	*0.70* [b]	*0.66*	*0.61*
RMSE[c]	0.66	0.96	1.00	*0.86*	*0.81*	*0.73*
SMAPE[d]	7.64	11.57	12.10	*10.18*	*9.64*	*8.64*
Spearman	0.86	0.68	0.62	*0.72*	*0.76*	*0.79*
Pearson	0.85	0.64	0.59	*0.74*	*0.76*	*0.81*
Obesity						
MAE	1.31	1.68	1.74	*1.57*	1.59	1.81
RMSE	1.66	2.56	2.80	*2.02*	*1.87*	2.45
SMAPE	5.81	7.31	8.05	*7.24*	*6.92*	7.88
Spearman	0.89	0.82	*0.85*	*0.87*	*0.89*	0.85
Pearson	0.95	0.86	0.86	*0.95*	*0.95*	0.91

[a]MAE: mean absolute error.

[b]The values in *italics* signify an improvement in the performance in comparison to the previous method.

[c]RMSE: root mean squared error.

[d]SMAPE: symmetric mean absolute percentage error.

Table 5. Results for the transfer learning framework across the 2 target variables for Canada for generalizability of the method trained over the years 2008-2014, and tested over the years 2016-2018.

Target variable	Trivial Baseline	Spatial	Spatio-temporal model	Multivariate spatio-temporal model	Lagged multivariate spatio-temporal model	Hierarchical lagged multivariate spatio-temporal model
Diabetes						
MAE[a]	0.84	0.82	*0.75* [b,c]	*0.71* [c]	*0.68* [c]	0.74[c]
RMSE[d]	1.07	*1.04* [c]	*0.93* [c]	0.93[c]	*0.86* [c]	0.99[c]
SMAPE[e]	11.16	*10.74* [c]	*9.78* [c]	*9.37* [c]	*9.00* [c]	10.05[c]
Spearman	0.69	*0.82* [c]	*0.84* [c]	0.76[c]	*0.78* [c]	0.78[c]
Pearson	0.7	*0.80* [c]	*0.82* [c]	0.76[c]	*0.77* [c]	*0.78* [c]
Obesity						
MAE	1.57	7.98	8.72	*3.39*	*2.62*	5.69
RMSE	2.37	8.4	9	*3.86*	*3.38*	5.92
SMAPE	4.99	29.59	33.59	*11.73*	*8.36*	20.71
Spearman	0.93	0.86	*0.89*	*0.92*	*0.93*	*0.95* [c]
Pearson	0.91	0.88	*0.9*	*0.93* [c]	0.91	*0.95* [c]

[a]MAE: mean absolute error.

[b]The values in *italics* signify an improvement in the performance in comparison to the previous method.

[c]The method beat the trivial baseline.

[d]RMSE: root mean squared error.

[e]SMAPE: symmetric mean absolute percentage error.

Table 6. The results for the transfer learning framework across the 2 target variables for Canada for generalizability of the US trained model.

Cross-country generalizability of the US-based model	Diabetes		Obesity	
	Trivial baseline	Lagged multivariate model	Trivial baseline	Lagged multivariate model
Train years: 2005-2018 (US); test years: 2008-2014 (Canada)				
MAE[a]	0.68	0.88	1.53	1.78
RMSE[b]	0.92	1.10	1.96	2.16
SMAPE[c]	9.9	12.65	6.91	8.20
Spearman	0.81	0.77	0.90	0.90
Pearson	0.76	0.74	0.91	0.91
Train years: 2005-2018 (US); test years: 2016-2018 (Canada)				
MAE	0.84	1.29	1.57	*1.54*
RMSE	1.07	1.49	2.37	*2.25*
SMAPE	11.16	16.04	4.99	*4.92*
Spearman	0.69	0.59	0.93	0.92
Pearson	0.70	0.60	0.91	0.91
Train years: 2005-2018 (US); test years: 2013-2014 (Canada)				
MAE	0.54	0.91	1.31	1.31
RMSE	0.66	1.12	1.66	*1.60* [d]
SMAPE	7.64	13.10	5.81	*5.76*
Spearman	0.86	0.74	0.89	*0.90*
Pearson	0.85	0.74	0.95	0.95

[a]MAE: mean absolute error.

[b]RMSE: root mean squared error.

[c]SMAPE: symmetric mean absolute percentage error.

[d]The values in italics signify improvement in performance over the trivial baseline.

Discussion

Principal Findings and Contributions

Value and Validity of Modeling Noncommunicable Diseases

The *slow-moving* nature of NCDs compared with the relatively faster moving trends in search behavior makes it a hard problem to perform lifestyle disease surveillance using Google Trends. The background literature reports overly optimistic results in this arena. This is potentially a consequence of what is known as the *positive result bias* [67] in the scientific community, leading to claims that *most* published research findings may be false [68,69]. We show in this work that in addition to the methodological shortcomings, *none* of the previous studies compares its results to the trivial *last year equals this year* baselines, which, surprisingly, is hard to beat. In this study, we empirically test the feasibility of the task by experimenting across different methods for 3 target variables. Our experiments are ordinal in nature as each subsequent experiment is an extension of the previous one. Although most of our latter extensions beat the trivial baseline, modeling NCDs is not a trivial problem. The main challenge of the problem lies in the scarcity of ground truth data, which is typically only available on an annual basis. Even if the data were available on a finer granularity, the inherent nature of NCDs, such as diabetes and obesity, does not allow for discernible monthly or weekly variation. Even the relative year-to-year changes are low which is why the correlation coefficients for the trivial baselines are high, and hard to beat. As a result, we anticipate a low-to-moderate value in modeling the *estimation* of NCDs, as well as in the validity of Google Trends for nowcasting lifestyle diseases.

The lack of validity of Google Trends in the given context can be partly attributed to changes in Web search behavior across time. As an example, Web search users might long have realized the potential to use Google for navigational queries, instead of having to remember and type exact website URLs. However, the use for informational queries in the health domains is likely to be still growing, also as Google adds new features, trying to answer common health-related questions directly on the search result page. A related but similar reason is that the meaning and perception of terms themselves can change over time. In linguistics, this phenomenon is known as the *semantic shift* [70] describing how the senses of words drift over time. A typical example of that is the word *Gay* which evolved from its meaning of *lighthearted* and *joyous* in the 1900s to *homosexual* in the 1990s [71]. With increasing popularity of social media sites and

Web-based tools, Web-based content is being produced expeditiously, leading to relatively frequent semantic shifts. Accounting for these complex phenomena with the scarcity of data is already a difficult problem. Additional limitations of the Google Trends framework make it an even challenging issue. These limitations include the changes in geolocation assignment applied to Google Trends in 2011, Google Trends' normalization scheme, and, finally, instability of search indices of a given keyword on different days [31,57].

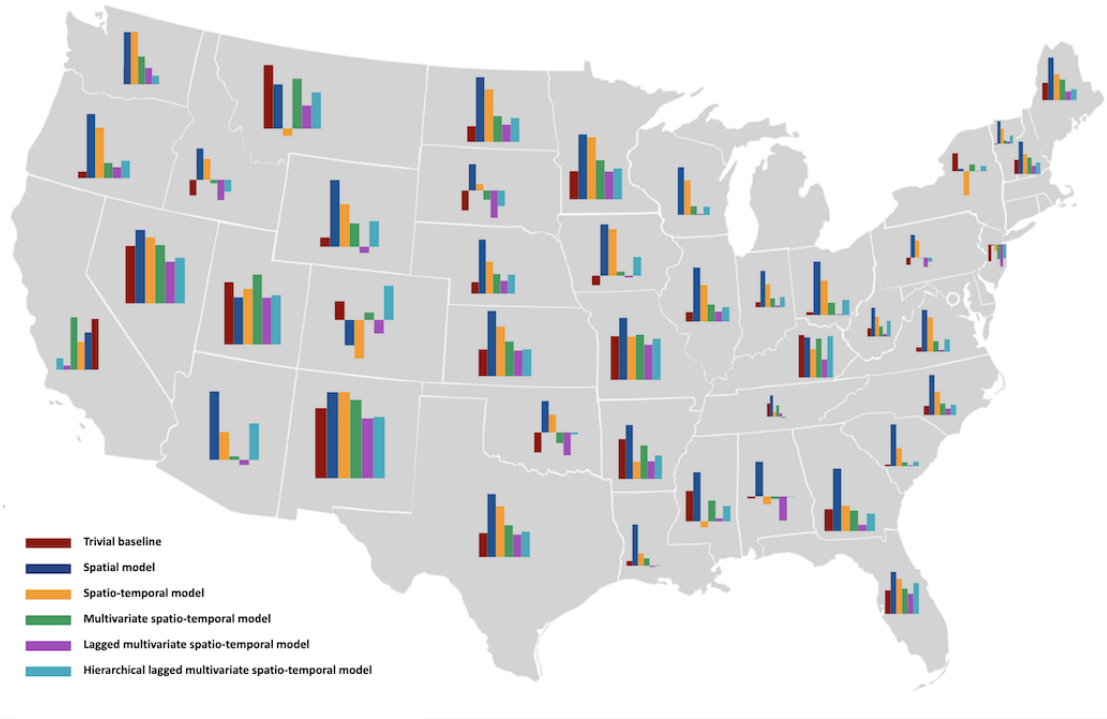
As for the first limitation, in 2011, Google implemented significant improvements in the geolocation of search queries (a note on the Web interface for Google Trends says: "An improvement to our geographical assignment was applied from 1/1/2011.") To account for these changes, in the earlier set of experiments for this study (not reported in this manuscript), we limited our training set to data from 2011 onward only. Unfortunately, none of our experimental models beat the trivial baseline. In the current set of experiments, we include the years from before 2011 to boost the performance, assuming that the geolocation assignment at the state level was not affected by the changes. We observe that this, in fact, is true and that more data helps learn better models. On the surface, this is not an interesting finding as machine learning models are inherently data hungry. However, we hypothesize that the boost in performance may also be attributed to learning semantic and usage shifts in data. By including a wider time window, we believe our models may implicitly be selecting features that are shift-independent, and pruning out features that are not. For example, the keyword *slim*, which was part of our obesity-related models, was assigned a weight of zero. Since Google Trends provides related topics for each keyword, entering the keyword *slim* yields topics such as *slim-fit pants* and *Plexus* along with *Xbox-Console* and *PlayStation 3*. This indicates that the keyword *slim* is used in 2 different contexts: obesity and slim console games. There was a potential semantic shift after 2010 when Xbox 360 slim model was released [72], and it is possible that our model was able to capture it given a wider time window, resulting in pruning it out.

A final limitation of Google Trends in the context of this study is the instability of search indices. We observe that repeatedly asking for the same data from Google Trends' can return different results. To correct for this, like [31], we resolve potential data instability issues by averaging data points collected across 10 consecutive days for the United States, and across 3 consecutive days for Canada.

Model Selection and Learning

In our experiments, we observe that a time-lagged multivariate model, which uses both ground truth and search data from the past year to nowcast the current year, improves over our previous, simpler methods. This suggests that the implications of the search behavior are observed in the NCD prevalence of later year rather than the immediate one. A final extension to our models is the use of state-ID as a covariate to model a hierarchical distribution. The inclusion of these state-level offsets significantly improves the results for diabetes and exercise. One interesting observation is that almost all of our models seem to underestimate the rate of NCDs. This is apparent in Figure 2, which shows how the models consistently underestimate the prevalence of obesity. We observe the same behavior across the other target variables. A potential reason is that of users moving away from the keywords used in this study as a consequence of a change in search behavior. This ties back to the discussion on the semantic and search behavior changes of users across time.

Figure 2. The comparison of the errors of 6 different methods for the target variable obesity for the year 2018 for each state. (Note: The bars represent the simple error [ie, ground truth prediction] for each state, and not the predicted diabetes or obesity rates. Therefore, the height of the bars is only comparable within each state and not comparable across states as the scale is not fixed. As most bars are above zero, this indicates that in most cases, the models underestimate the ground truth obesity rates.).



Geographical Transfer Learning

This work is one of the very few to test the geographical generalizability of the proposed method. We use Canada as a test case as it is close to the United States, and is one of the very few English-speaking countries to have reliable *state-level* statistics. On the basis of our evaluation, most of the models trained on Canada do not beat the baseline. This can again be attributed to the scarcity of data as in comparison with the United States, Canada has only 10 provinces and the usable data are only available starting 2007. Despite the failure to beat the baseline, we see opportunities for public health monitoring in countries where the public health monitoring system is less developed and so the otherwise trivial baseline cannot be applied. In particular, it seems feasible to monitor relative across-state or across-year trends, rather than absolute prevalence rates. Extending our approach to countries with low resources comes with certain challenges though as, exactly due to their lack of trustworthy data, it is hard to have an objective evaluation of the system performance.

Proactivity Versus Reactivity

In terms of the health-related behavior, the user search behavior can be weakly classified into 2 categories: proactive and reactive. Proactive search behavior is when the search is derived from curiosity and awareness. This usually defines users who are cautious about their health with an interest in preventing a health condition. Reactive, on the other hand, defines behavior, which is corrective where users are seeking help to treat or manage their health problems. In the case of diabetes and obesity in the United States and Canada, we observe that predictive search terms indicate a highly reactive behavior, with *diabetes symptoms*, *diabetic diet, and exercise* carrying higher absolute weights (and, hence, higher significance). This potentially indicates that it is mostly users affected by a condition who show interest. If this interpretation holds, this points toward a need to carry out public health interventions to try to change the information seeking to be more proactive.

Keyword Selection Using Google Correlate

For our study, we tried to limit the risk of overfitting by following a robust keyword selection process both in the preprocessing phase as well as the modeling phase. As a preprocessing step, we used Semantic-Link, Google Trends–related keywords, and Google Correlate as tools to select the best set of features. This is one of the key strengths of our study where we combine semantically connected, cooccurring, and correlated terms to create a diverse keyword list. In the context of NCDs, this is the first study to utilize the strength of Google Correlate in the process of keyword selection. An interesting case study in this context is that of obesity-related keywords, a subset of which is shown in Table 2. One of the top keywords is *dresses plus size*, which is also included in the L1 regularized model. Although it is a fairly intuitive keyword in hindsight, it would have been challenging to discover it without Google Correlate. Although we do not perform an in-depth analysis of the keywords in this study, we encourage researchers to use pseudo bootstrapping approaches in their keyword selection process to perform richer analysis. In terms of Google Correlate, unfortunately, a note on the Web interface says: “Google Correlate will shut down on December 15th 2019 as a result of low usage,” as a result of which it will not be available for future researchers.

Limitations

Currently, our approach is agnostic to word semantics and we just explore what is predictive of a particular NCD without accounting for anomalies. In this context, recent methods on the use of word embeddings may be useful to curb inconsistent patterns of associations. It would also be useful if we can explicitly capture the semantic or behavior shifts in time, and make it part of the model. This warrants use of sophisticated dynamic network analysis tools and other time-series models.

One more potential limitation of this study is the inability to test our methodology on other countries. Although we show promising results for Canada, it would have been preferable to test our methods to other English, and even non-English–speaking countries. This limitation was primarily a consequence of the unavailability of state-level ground truth for many countries.

Finally, one limitation pertaining to our data collection process is that while like [31], we resolve potential data instability issues by averaging data points collected across 10 consecutive days for the United States, we use only 3 data points for Canada. We do this in the interest of time, as the experiments for Canada are more preliminary. Nonetheless, to get a sense of whether variation between days might affect the overall conclusions, we conducted a day-to-day test-retest reliability analysis across 3 days from Google Trends data for Canada, using Spearman and Pearson correlation. We computed the correlation coefficient between each set of data collected on day 1 and day 2, and between data collected on day 2 and day 3. For day 1 and day 2, we found an average Spearman and Pearson correlation of 0.85 and 0.86 for temporal data, and of 0.84 and 0.88 for the spatial data. For day 2 and day 3, we found an average Spearman and Pearson correlation of 0.85 and 0.86 for temporal data, and of 0.84 and 0.89 for the spatial data. Given that (1) the day-to-day correlations are very high, and that (2) we nevertheless average across 3 days, we do not expect our conclusions for Canada to be affected by using data points from only 3 rather than 10 days.

Future Research Directions

Using Google Trends for *nowcasting or estimating* the spatio-temporal prevalence of NCDs may not always be of significant value due to the scarcity of data, uncertainty of search behavior, and the changes in usage of different keywords across time. However, there are still several fronts that remain unexplored.

In terms of the NCDs, using Google Trends for *qualitative* research is an overlooked territory. In this context, feature analysis is a promising venue where tracking, investigating, and discovering different keywords could help policy makers or governments make better decisions.

In terms of the qualitative analysis, it might also be useful to employ Google Trends for analyzing NCD-related *events*. In this context, events could be described as short-term (or single-day) phenomenon. A popular example of that is initiating and monitoring public health campaigns and interventions. For health campaigns, it is useful to conduct both prehoc and posthoc analyses. Prehoc analyses include determining the target audience, and baseline search behaviors of people. On the other hand, posthoc analyses characterize changes in the search behavior triggered by the target event. Examples of related work in the context of public health awareness campaigns can be found in [73-76].

Another promising future direction includes the use of task-specific modeling techniques where the domain knowledge about the task at hand could be used to craft better machine learning models. In our study, we use the state-ID as a proxy to model hierarchical distribution with different country-level and state-level behaviors. However, our models are still rudimentary and one could potentially employ more sophisticated models, such as hierarchical linear models [77], or neural regression trees [78], to model different hierarchies. In our study, we model the hierarchy based on the state-ID. The state, in this case, is treated as a *cluster* selected a priori to the modeling based on the domain knowledge. One could also try to find optimal clusters (such as group of similar states) or partitioning of the data simultaneously while modeling the problem itself. This can be achieved by using techniques such as nearest neighbors [79] or other tree-based approaches such as those shown in [78,80]. An added advantage of the hierarchical models is their increased interpretability resulting in some qualitative insights. Such models also allow us to optimize feature selection on subnational levels while modeling a national distribution, improving the overall robustness. Clustering approaches could also lead to *directed* public health interventions by providing cluster-specific insights. Governmental public health interventions would be much more effective if they target the right audience, and in a right and directed manner, taking into account the local search behavior.

In summary, we believe that despite limited promise for quantitative NCD surveillance, Google Trends still holds value in relation to NCDs, particularly for qualitative analyses or for monitoring the effect of public health campaigns. Finally, in terms of other fast-moving trends, any useful spatio-temporal national-level surveillance using Google Trends necessitates the use of appropriate denormalization such as the one shown in this work.

Conclusion

In this paper, we first review the methods presented in the background literature. We present a comprehensive table where we compare the literature across different metrics. One of the surprising findings of our study is the absence of evaluation against trivial baselines in all of the reviewed papers. We furthermore, highlight the methodological shortcomings of the most relevant research papers. In the second part of the paper, we use a corrective approach to improve upon the background work. Specifically, we explore the feasibility of using Google Trends for nowcasting the prevalence of lifestyle diseases in the context of diabetes, obesity, and exercise. To undo the effect of Google Trends' normalization, we propose a novel spatio-temporal denormalization scheme. That combined with the trivial baseline as a covariate beat the previously set baseline methods in the background literature for most of the target variables. We further improve upon that model by shifting the time window by 1 year, and then including a state-ID as a

covariate. Our best models beat the trivial baseline with a significant improvement in performance. We, however, realize that this requires both abundant data, as well as creative modeling strategies. Furthermore, we extend upon our formative work to show generalizability of our methodology and trained models in the international setting, setting a cornerstone for using such transfer learning-based approaches in low-resource countries. Finally, we propose various possible future paths researchers can take to conduct both quantitative and qualitative analyses using Google Trends.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The final postpruning keywords used in the modeling of each of the different target variables.
[DOCX File , 9 KB - jmir_v22i1e13347_app1.docx]

Multimedia Appendix 2
Supplementary information on Google Trends' denormalization procedure.
[PDF File (Adobe PDF File), 1461 KB - jmir_v22i1e13347_app2.pdf]

References

1. Thacker S, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev 1988;10:164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
2. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
3. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
4. Granka L. Using online search traffic to predict US presidential elections. PS Polit Sci Polit 2013;46(2):271-279. [doi: 10.1017/s1049096513000292]
5. Ojala J, Zagheni E, Billari FC, Weber I. Fertility and its Meaning: Evidence from Search Behavior. In: Proceedings of the Eleventh International AAAI Conference on Web and Social Media. 2017 Presented at: ICWSM'17; May 15-18, 2017; Montréal, Québec, Canada p. 640-643.
6. Preis T, Moat HS, Stanley HE. Quantifying trading behavior in financial markets using Google Trends. Sci Rep 2013;3:1684 [FREE Full text] [doi: 10.1038/srep01684] [Medline: 23619126]
7. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with web search. Proc Natl Acad Sci U S A 2010 Oct 12;107(41):17486-17490 [FREE Full text] [doi: 10.1073/pnas.1005962107] [Medline: 20876140]
8. Choi H, Varian H. Predicting the present with Google trends. Econ Rec 2012;88(1):2-9. [doi: 10.1111/j.1475-4932.2012.00809.x]
9. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
10. Milinovich GJ, Avril SM, Clements AC, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. BMC Infect Dis 2014 Dec 31;14:690 [FREE Full text] [doi: 10.1186/s12879-014-0690-1] [Medline: 25551277]
11. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
12. Johnson AK, Mehta SD. A comparison of internet search trends and sexually transmitted infection rates using Google trends. Sex Transm Dis 2014 Jan;41(1):61-63. [doi: 10.1097/OLQ.0000000000000065] [Medline: 24326584]
13. Polgreen P, Chen Y, Pennock D, Nelson F. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
14. World Health Organization. 2018 Jun 1. Noncommunicable Diseases URL: http://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2018-11-03] [WebCite Cache ID 73eOmOdLJ]
15. Healthy Communities for a Healthy Future. Economic Costs of Obesity URL: http://www.healthycommunitieshealthyfuture.org/learn-the-facts/economic-costs-of-obesity/ [accessed 2019-12-24]
16. Burns SM, Turner DP, Sexton KE, Deng H, Houle TT. Using search engines to investigate shared migraine experiences. Headache 2017 Sep;57(8):1217-1227 [FREE Full text] [doi: 10.1111/head.13130] [Medline: 28660638]
17. Parker J, Cuthbertson C, Loveridge S, Skidmore M, Dyar W. Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. J Affect Disord 2017 Apr 15;213:9-15. [doi: 10.1016/j.jad.2016.10.038] [Medline: 28171770]

18. McCarthy MJ. Internet monitoring of suicide risk in the population. J Affect Disord 2010 May;122(3):277-279 [FREE Full text] [doi: 10.1016/j.jad.2009.08.015] [Medline: 19748681]
19. Ma-Kellams C, Or F, Baek JH, Kawachi I. Rethinking suicide surveillance: Google search data and self-reported suicidality differentially estimate completed suicide risk. Clin Psychol Sci 2016;4(3):480-484. [doi: 10.1177/2167702615593475]
20. Chandler V. Google and suicides: what can we learn about the use of internet to prevent suicides? Public Health 2018 Jan;154:144-150. [doi: 10.1016/j.puhe.2017.10.016] [Medline: 29245021]
21. Gunn JF, Lester D. Using google searches on the internet to monitor suicidal behavior. J Affect Disord 2013 Jun;148(2-3):411-412. [doi: 10.1016/j.jad.2012.11.004] [Medline: 23182592]
22. Bragazzi NL. A Google Trends-based approach for monitoring NSSI. Psychol Res Behav Manag 2013;7:1-8 [FREE Full text] [doi: 10.2147/PRBM.S44084] [Medline: 24376364]
23. Yang AC, Tsai S, Huang NE, Peng C. Association of internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord 2011 Jul;132(1-2):179-184. [doi: 10.1016/j.jad.2011.01.019] [Medline: 21371755]
24. Hagihara A, Miyazaki S, Abe T. Internet suicide searches and the incidence of suicide in young people in Japan. Eur Arch Psychiatry Clin Neurosci 2012 Feb;262(1):39-46. [doi: 10.1007/s00406-011-0212-8] [Medline: 21505949]
25. Sueki H. Does the volume of internet searches using suicide-related search terms influence the suicide death rate: data from 2004 to 2009 in Japan. Psychiatry Clin Neurosci 2011 Jun;65(4):392-394 [FREE Full text] [doi: 10.1111/j.1440-1819.2011.02216.x] [Medline: 21569178]
26. Bruckner TA, McClure C, Kim Y. Google searches for suicide and risk of suicide. Psychiatr Serv 2014 Feb 1;65(2):271-272 [FREE Full text] [doi: 10.1176/appi.ps.201300211] [Medline: 24492910]
27. Arendt F. Suicide rates and information seeking via search engines: a cross-national correlational approach. Death Stud 2018 Sep;42(8):508-512. [doi: 10.1080/07481187.2017.1388305] [Medline: 29173051]
28. Yang AC, Huang NE, Peng C, Tsai S. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. PLoS One 2010 Oct 28;5(10):e13728 [FREE Full text] [doi: 10.1371/journal.pone.0013728] [Medline: 21060851]
29. Song TM, Song J, An J, Hayman LL, Woo J. Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. Yonsei Med J 2014 Jan;55(1):254-263 [FREE Full text] [doi: 10.3349/ymj.2014.55.1.254] [Medline: 24339315]
30. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
31. Tran US, Andel R, Niederkrotenthaler T, Till B, Ajdacic-Gross V, Voracek M. Low validity of Google Trends for behavioral forecasting of national suicide rates. PLoS One 2017;12(8):e0183149 [FREE Full text] [doi: 10.1371/journal.pone.0183149] [Medline: 28813490]
32. Phillips CA, Leahy AB, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: Retrospective study. J Med Internet Res 2018 Jan 8;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
33. Wehner MR, Nead KT, Linos E. Correlation among cancer incidence and mortality rates and internet searches in the United States. JAMA Dermatol 2017 Sep 1;153(9):911-914 [FREE Full text] [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]
34. Ricketts CF, Silva CG. An analysis of morbidity and mortality using Google Trends. J Hum Behav Soc Environ 2017 Mar 28;27(6):559-570. [doi: 10.1080/10911359.2017.1299658]
35. Sarigul S, Rui H. Nowcasting Obesity in the US Using Google Search Volume Data. In: Proceedings of the Joint Symposium: Social Networks, Social Media and the Economics of Food. 2014 Presented at: AAEA/EAAE/CAES'14; May 29-30, 2014; Montreal, Canada URL: http://ageconsearch.umn.edu/record/166113 [doi: 10.22004/ag.econ.166113]
36. Breyer BN, Eisenberg ML. Use of Google in study of noninfectious medical conditions. Epidemiology 2010 Jul;21(4):584-585 [FREE Full text] [doi: 10.1097/EDE.0b013e3181e080cd] [Medline: 20539114]
37. Aguirre PE, Coelho M, Oliveira T, Rios D, Cruvinel AF, Cruvinel T. What can Google inform us about people's interests regarding dental caries in different populations? Caries Res 2018;52(3):177-188. [doi: 10.1159/000485107] [Medline: 29353276]
38. Lotto M, Aguirre PE, Rios D, Machado MA, Cruvinel AF, Cruvinel T. Analysis of the interests of Google users on toothache information. PLoS One 2017;12(10):e0186059 [FREE Full text] [doi: 10.1371/journal.pone.0186059] [Medline: 29049315]
39. Cruvinel T, Aguirre PE, Lotto M, Oliveira TM, Rios D, Cruvinel AF. Digital behavior surveillance: monitoring dental caries and toothache interests of Google users from developing countries. Oral Dis 2019 Jan;25(1):339-347. [doi: 10.1111/odi.12986] [Medline: 30270556]
40. Leffler CT, Davenport B, Chan D. Frequency and seasonal variation of ophthalmology-related internet searches. Can J Ophthalmol 2010 Jun;45(3):274-279. [doi: 10.3129/i10-022] [Medline: 20436544]
41. Walcott BP, Nahed BV, Kahle KT, Redjal N, Coumans J. Determination of geographic variance in stroke prevalence using internet search engine analytics. Neurosurg Focus 2011 Jun;30(6):E19. [doi: 10.3171/2011.2.FOCUS1124] [Medline: 21631220]

42. Nguyen T, Tran T, Luo W, Gupta S, Rana S, Phung D, et al. Web search activity data accurately predict population chronic disease risk in the USA. J Epidemiol Community Health 2015 Jul;69(7):693-699. [doi: 10.1136/jech-2014-204523] [Medline: 25805603]
43. Ingram DG, Plante DT. Seasonal trends in restless legs symptomatology: evidence from internet search query data. Sleep Med 2013 Dec;14(12):1364-1368. [doi: 10.1016/j.sleep.2013.06.016] [Medline: 24152798]
44. Braun T, Harréus U. Medical nowcasting using Google Trends: application in otolaryngology. Eur Arch Otorhinolaryngol 2013 Jul;270(7):2157-2160. [doi: 10.1007/s00405-013-2532-y] [Medline: 23632877]
45. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. Urology 2013 Jan;81(1):37-42. [doi: 10.1016/j.urology.2011.04.024] [Medline: 21676450]
46. Bragazzi NL. Infodemiology and infoveillance of multiple sclerosis in Italy. Mult Scler Int 2013;2013:924029 [FREE Full text] [doi: 10.1155/2013/924029] [Medline: 24027636]
47. Brigo F, Igwe SC, Ausserer H, Nardone R, Tezzon F, Bongiovanni LG, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. Epilepsy Behav 2014 Feb;31:67-70. [doi: 10.1016/j.yebeh.2013.11.020] [Medline: 24361764]
48. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with Google trends: evidence from Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
49. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]
50. Hassid BG, Day LW, Awad MA, Sewell JL, Osterberg EC, Breyer BN. Using search engine query data to explore the epidemiology of common gastrointestinal symptoms. Dig Dis Sci 2017 Mar;62(3):588-592. [doi: 10.1007/s10620-016-4384-y] [Medline: 27878646]
51. Coogan S, Sui Z, Raubenheimer D. Gluttony and guilt: monthly trends in internet search query data are comparable with national-level energy intake and dieting behavior. Palgrave Commun 2018;4(1):-. [doi: 10.1057/s41599-017-0055-7]
52. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System URL: https://www.cdc.gov/brfss/ [accessed 2019-12-24]
53. Pacula M. Semantic Link. 2011. URL: http://semantic-link.com/ [accessed 2019-12-24]
54. Lampos V, Zou B, Cox IJ. Enhancing Feature Selection Using Word Embeddings: The Case of Flu Surveillance. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 695-704. [doi: 10.1145/3038912.3052622]
55. Google Correlate: FAQ URL: https://www.webcitation.org/73eSmswKp [accessed 2018-11-03] [WebCite Cache ID 73eSmswKp]
56. Github. PyTrends URL: https://github.com/GeneralMills/pytrends [accessed 2018-11-03] [WebCite Cache ID 73eSitocg]
57. Carrière-Swallow Y, Labbé F. Nowcasting with Google Trends in an emerging market. J Forecast 2013;32(4):289-298. [doi: 10.1002/for.1252]
58. Google Help.: Google Compare Trends search terms URL: https://support.google.com/trends/answer/4359550?ref_topic=4365530 [accessed 2018-11-03] [WebCite Cache ID 73eSdylT7]
59. Google Help. FAQ about Google Trends data URL: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052 [accessed 2019-12-24] [WebCite Cache ID 73eSYBJXk]
60. Memon SA, Razak S, Weber I. LinkedIn SlideShare. 2019 Sep 23. Google Trends Normalization and Denormalization URL: https://www.slideshare.net/ShahanAliMemon/google-trends-denormalized-175072435/ShahanAliMemon/google-trends-denormalized-175072435 [accessed 2019-12-24]
61. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
62. Scikit-learn: Machine learning in Python. URL: https://scikit-learn.org/stable/ [accessed 2019-12-24]
63. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
64. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 1970;12(1):55-67. [doi: 10.1080/00401706.1970.10488634]
65. Makridakis S. Accuracy measures: theoretical and practical concerns. Int J Forecast 1993;9(4):527-529. [doi: 10.1016/0169-2070(93)90079-3]
66. Statistics Canada. URL: https://www.statcan.gc.ca/eng/start [accessed 2019-09-20]
67. Plüddemann A, Banerjee A, O'Sullivan J. Catalogue Of Biases. 2017. Positive results bias URL: https://catalogofbias.org/biases/positive-results-bias/ [accessed 2019-12-24]
68. Ioannidis JP. Why most published research findings are false. PLoS Med 2005 Aug;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
69. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci 2014 Nov;1(3):140216 [FREE Full text] [doi: 10.1098/rsos.140216] [Medline: 26064558]

70. Wikipedia, the free encyclopedia. Semantic change URL: https://en.wikipedia.org/wiki/Semantic_change [accessed 2019-12-24]
71. Hamilton W, Leskovec J, Jurafsky D. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016 Presented at: EMNLP'16; November 1–5, 2016; Austin, Texas, USA p. 2116-2121 URL: http://europepmc.org/abstract/MED/28580459 [doi: 10.18653/v1/d16-1229]
72. Strauss P. Technabob. 2010 Jun 24. New Xbox 360 Slim Price, Release Date and Specs Revealed URL: https://technabob.com/blog/2010/06/14/new-xbox-360-slim-price-release-date-specs/ [accessed 2019-12-24]
73. Shariatpanahi SP, Jafari A, Sadeghipour M, Azadeh-Fard N, Majidzadeh-A K, Farahmand L, et al. Assessing the effectiveness of disease awareness programs: evidence from Google Trends data for the world awareness dates. Telematics Informatics 2017;34(7):904-913. [doi: 10.1016/j.tele.2017.03.007]
74. Ortiz-Martinez Y, Rios-González CM. Global impact of the World Hepatitis Day 2016: an evaluation using Google trends. J Infect Public Health 2017;10(5):690-691 [FREE Full text] [doi: 10.1016/j.jiph.2016.09.015] [Medline: 28169192]
75. Ayers JW, Althouse BM, Allem J, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012 May 28;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]
76. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
77. Calvin JA, Bryk AS, Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Technometrics 1994;36(1):116-117. [doi: 10.2307/1269211]
78. Memon SA, Zhao W, Raj B, Singh R. Neural Regression Trees. In: Proceedings of the 2019 International Joint Conference on Neural Networks.: IEEE; 2019 Presented at: IJCNN'19; July 14-19, 2019; Budapest, Hungary URL: https://arxiv.org/abs/1810.00974 [doi: 10.1109/ijcnn.2019.8852133]
79. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21-27. [doi: 10.1109/TIT.1967.1053964]
80. Zhao W, Gao Y, Memon SA, Raj B, Singh R. arXiv e-Print archive. 2019. Hierarchical Routing Mixture of Experts URL: https://arxiv.org/abs/1903.07756 [accessed 2019-12-24]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
GFT: Google Flu Trend
ILI: influenza-like illness
MAE: mean absolute error
NCDs: noncommunicable diseases
RMSE: root mean squared error
SMAPE: symmetric mean absolute percentage error
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 09.01.19; peer-reviewed by U Tran, P Ayala Aguirre; comments to author 27.04.19; revised version received 08.10.19; accepted 29.11.19; published 27.01.20.

Please cite as:
Memon SA, Razak S, Weber I
Lifestyle Disease Surveillance Using Population Search Behavior: Feasibility Study
J Med Internet Res 2020;22(1):e13347
URL: http://www.jmir.org/2020/1/e13347/
doi:10.2196/13347
PMID:32012050

Original Paper

Prevalence and Outcomes of Web-Based Health Information Seeking for Acute Symptoms: Cross-Sectional Study

Lydia Aoun[1,2], MD; Najla Lakkis[1], MD, MSc; Jumana Antoun[1], MD, MSc

[1]Department of Family Medicine, American University of Beirut, Beirut, Lebanon

[2]Faculty of Medicine, Beirut Arab University, Beirut, Lebanon

Corresponding Author:
Jumana Antoun, MD, MSc
Department of Family Medicine
American University of Beirut
Beirut
Lebanon
Phone: 961 1350000 ext 3049
Email: ja46@aub.edu.lb

Abstract

Background: The literature indicates that Web-based health information seeking is mostly used for seeking information on well-established diseases. However, only a few studies report health information seeking in the absence of a doctor's visit and in the context of acute symptoms.

Objective: This survey aimed to estimate the prevalence of Web-based health information seeking for acute symptoms and the impact of such information on symptom management and health service utilization.

Methods: This was a cross-sectional study of a convenience sample of 287 Lebanese adults (with a response rate of 18.5% [54/291]) conducted between December 2016 and June 2017. The survey was answered by participants online or through phone-based interviews.

Results: A total of 64.3% of the participants (178/277) reported checking the internet for health information when they had an acute symptom. The rate of those who sought to use Web-based health information first when experiencing acute symptom(s) in the past 12 months was 19.2% (25/130). In addition, 50% (9/18) visited the doctor because of the obtained information, and the rest self-medicated or sought a pharmacist's advice; the majority (18/24, 75%) improved within 3-4 days.

Conclusions: Higher education level and trust in Web-based medical information were two major predictors of Web-based health information seeking for acute symptoms. Seeking Web-based health information first for acute symptoms is common and may lead to self-management by avoiding a visit to the physician. Physicians should encourage their patients to discuss Web-based health information and guide them toward trusted online websites.

(J Med Internet Res 2020;22(1):e15148) doi:10.2196/15148

KEYWORDS

internet; health information; acute symptoms; acute disease

Introduction

Background

Since the launch of the World Wide Web in 1990, the number of internet users has increased remarkably (49.6% worldwide [1] and 75.9% in Lebanon [2]) as did the number of Web-based health information seekers [3]. Most internet users visit online websites to find information about physical rather than mental illnesses [4-6] and those mainly related to a diagnosis of a new health problem, an ongoing medical condition, or a prescribed medication [5]. Moreover, most users seek Web-based health information before or after a doctor's visit [5] to improve their involvement in decision making or supplement information provided by their physicians [7]. They are motivated by the desire for reassurance or a second opinion to challenge other information and to improve their understanding of a condition [8].

Web-based health information seeking is common, and the consumers considered it beneficial. The internet has empowered patients with chronic conditions such as diabetes and cancer [9] and helped patients in making decisions about bariatric surgeries [10] and oncologic management [11]. From the perspective of

Web-based health information seekers, the health information they sought was useful [4], had a positive impact on their health [12], and improved their medical information [12] and self-care [7].

Study Objectives

Most of the literature indicated that Web-based health information seeking is mostly used for well-established diseases. However, only few studies reported health information seeking in the absence of a doctor's visit [5,7,13,14] and specifically in the context of acute symptoms (symptoms with an abrupt onset and usually a short course, eg, fever, back pain, headache, diarrhea, flu, and urinary frequency). Seeking information on the internet for acute complaints may lead to self-diagnosis and self-medication, which may result in a delay in treatment and incorrect choice of therapies [15,16]. This is more challenging in the context of Web-based health information, as the quality of the information of the websites is not standardized [17]. Therefore, it is worth studying the role of Web-based health information seeking in self-diagnosis of acute symptoms and the consequences of consulting Web-based health information. This survey aimed to estimate the prevalence of Web-based health information seekers for acute symptoms and the impact of such information on symptom management and health service utilization.

Methods

Approval

This cross-sectional survey–based study was conducted between December 2015 and June 2016 to identify the proportion of the public who sought to use the internet for information about acute symptoms. Ethical approval was granted by the institution review board of the American University of Beirut.

Participants and Sample Size

The target population was adults from the general public (aged ≥18 years) residing in Lebanon. The convenience sampling method was selected, and the recruitment method included Google Ads or phone-based interviews. Convenience sampling is widely used in social behavioral studies. As our target population comprised people who used Web-based health information, we elected to use an online recruitment method rather than traditional methods such as posters or inviting people from some geographic areas. In fact, there is growing interest in the literature on the use of Google Ads or Facebook ads to recruit people [18-20].

The sample size was calculated with the formula (Figure 1) used in prevalence studies [21,22]. In the absence of the literature reporting this prevalence, we expected the proportion of adults who use the internet for acute symptoms to be similar to that of adults who use the internet for health information, in general. Therefore, we set the expected population proportion as 80% with 95% CIs and the level of precision of estimate within 5%. The sample size required was 246 people.

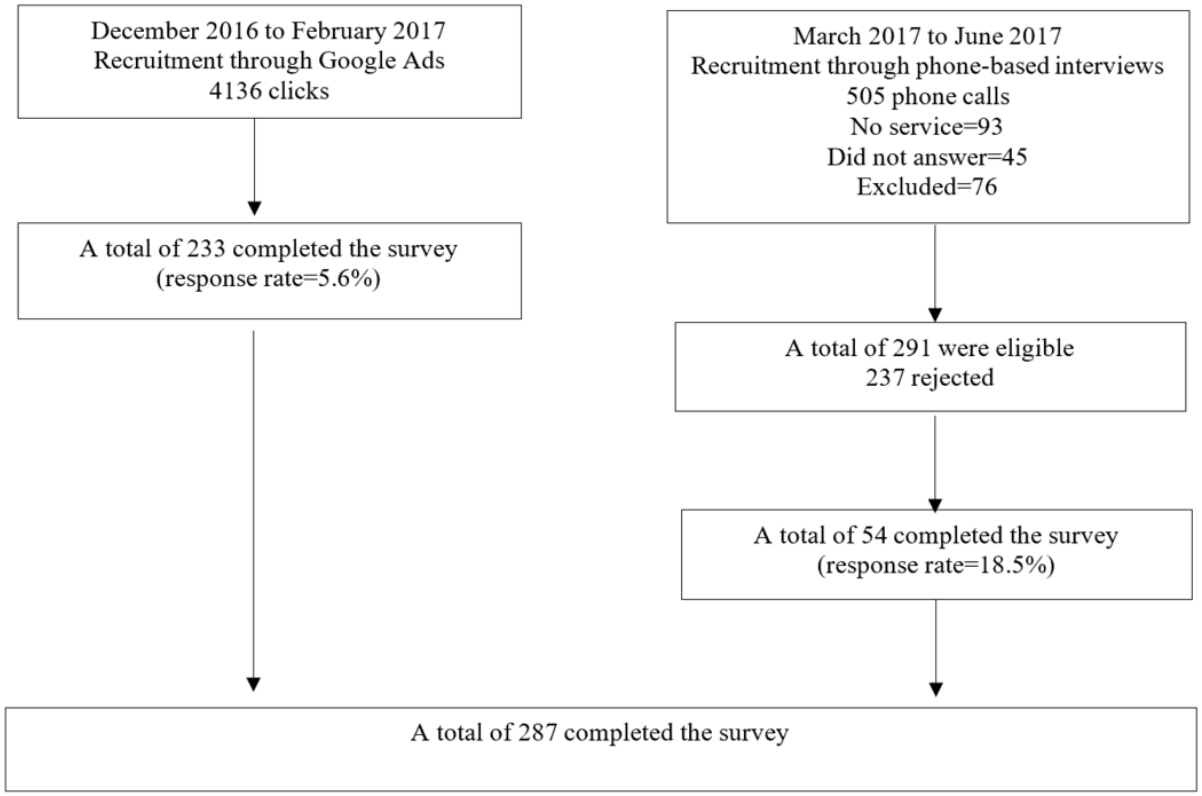
Figure 1. Sample size calculation.

$$n = \frac{(z^2)P(1-P)}{d^2}$$

Recruitment

A total of 233 participants were first recruited via Google Ads and were asked to complete a Web-based survey. After 3 months, we were not able to recruit enough sample size, so we changed the recruitment to phone-based interviews. The native language of the country is Arabic; however, the country is well known for its trilingual system (Arabic, English, and French). English is taught as a second language in private and public schools and is the language of communication in most universities [23]. For Google Ads, the survey was conducted in English; those who answered the Google Ads were more likely to know English because they were surfing the internet. For the phone-based interview recruitment, the questionnaire was translated into Arabic, the native language of the country, as we may have called people at their households. Using the RAND function in Microsoft Excel, a random list of phone numbers was generated; phone calls were made during fixed times of each day (day and night). Following this, an additional 54 participants were recruited. Figure 2 describes the process of recruiting participants.

Figure 2. Recruitment of the participants.



Survey Questions

The survey included information about participants' demographics, general internet use, and health information seeking. To measure the proportion of participants who sought the internet for information about an acute event, we used the following question: How often do you search the internet for information about an *acute* health event? (An acute illness is defined as a disease with an abrupt onset and usually a short course, eg, headache, fever, flu, pain, diarrhea, urinary frequency, and back pain). Participants were then asked if they had an acute symptom in the past 12 months, whether they sought Web-based health information regarding their acute symptom and their further actions (visit a doctor, pharmacy, or self-medicate), and what were the final illness outcomes (improved or got worse).

Statistical Analysis

A descriptive analysis was used to describe the demographics, access to Web-based health information, and outcomes. Means were used for continuous variables, and percentages were used for categorical variables. A bivariate analysis using the Chi-square test was used to investigate the relationship between Web-based health information seeking in the acute setting and patients' actions when they have an acute symptom. The participants' actions were also correlated with the demographics by using the Chi-square test, except for age, where the correlation was studied using the *t* test/one-way analysis of variance. The statistical significance was set at $P<.05$.

Results

Demographics

A total of 287 eligible participants completed the survey. The demographics of the participants are illustrated in Table 1. The mean age of the participants was 46.1 years; 16.7% (N=48) of them were older than 65 years and 6.8% were younger than 20 years. Almost two-thirds of the participants were married (n=139) and had university-level education (n=165), and 77% (n=193) of them had medical coverage. Around two-thirds of the participants did not have a chronic medical illness (n=196) and considered their health to be good (n=122) or very good (n=61).

Table 1. Demographics of the participants.

Characteristics	Values
Age (years), mean (SD)	46.1 (19.2)
Gender (N=242), n (%)[a]	
Male	118 (48.8)
Female	124 (51.2)
Marital status (N=242), n (%)[a]	
Single	82 (33.9)
Married	139 (57.4)
Divorced	9 (3.7)
Widowed	10 (4.1)
Separated	2 (0.8)
Educational status (N=240), n (%)[a]	
University	165 (57.5)
Technical training	29 (10.1)
Secondary school	29 (10.1)
Primary school	13 (4.5)
None	4 (1.4)
Monthly income (US $; N=222), n (%)[a]	
<500	43 (15.0)
500-1300	81 (28.2)
1300-3300	66 (23.0)
>3300	32 (11.1)
Medical coverage (N=252), n (%)[a]	
National Social Security Fund	78 (27.2)
Private insurance	101 (35.2)
Co-operative	10 (3.5)
None	59 (20.6)
Others	4 (1.4)
Chronic medical condition[a] (N=287), n (%)	91 (31.7)
Self-perceived health status (N=252), n (%)[a]	
Very good	61 (21.3)
Good	122 (42.5)
Moderate	52 (18.1)
Poor	14 (4.9)
Very poor	3 (1.0)
Have a physician whom they consult on regular basis (N=287), n (%)[a]	148 (51.6)
Frequency of doctors' visit in the past 1 year (N=255), n (%)[a]	
Never	49 (17.1)
1-2 times	123 (42.9)
3 times	32 (11.1)
≥4 times	51 (17.8)

[a]Missing values.

Internet Use for Health Information in General and for an Acute Health Event

Participants were provided with the definition of an acute illness: A disease with an abrupt onset and usually a short course, for example, headache, fever, flu, pain, diarrhea, urinary frequency, and back pain. Almost all the participants (235/249, 94.3%) reported general internet use at least twice weekly, whereas one-third of the participants (86/277, 31.0%) reported at least twice weekly internet use for an acute health event specifically (Table 2). One-third of the participants (97/249, 38.9%) rarely or never used the internet for health information, and only 10.3% of them (28/270) rarely or never trusted the health information obtained (Table 2).

Participants who used the internet for an acute illness were likely to be more educated (P=.01), have a doctor they regularly consult (P=.009), use the internet for health information (P≤.001), trust the Web-based health information found (P<.001), and have a middle monthly income (P=.044). It is also important to note that there was no significant relationship between searching the internet for an acute illness and gender, having a chronic medical condition, perceived health status, frequency of visits to the doctor, or general internet use.

Table 2. Internet use for health information in general and for an acute health event.

Characteristics	Value, n (%)
General internet use	
Daily	203 (70.7)
≥2 times/week	30 (10.5)
Once weekly	6 (2.1)
≥2 times/month	1 (0.30)
Once monthly	13 (4.5)
Rarely	6 (2.1)
Never	0 (0)
Health website use	
Daily	39 (13.6)
≥2 times/week	44 (15.3)
Once weekly	31 (10.8)
≥2 times/month	23 (8.0)
Once monthly	13 (4.5)
Rarely	68 (23.7)
Never	28 (9.8)
Internet search for acute health event[a]	
Daily	50 (17.4)
≥2 times/week	47 (17.3)
Once weekly	22 (7.7)
≥2 times/month	41 (14.3)
Once monthly	13 (4.5)
Rarely	79 (27.5)
Never	19 (6.6)
Trust in Web-based health information obtained	
Always	32 (11.1)
Very often	66 (23.0)
Sometimes	122 (42.5)
Rarely	19 (6.6)
Never	9 (3.1)
Not applicable	18 (6.3)

[a]An acute health event was defined as an acute illness with an abrupt onset and usually a short course (eg, headache, fever, sore throat, ear pain, cold/flu, diarrhea, urinary symptoms, and acute back pain).

Description of the Participants Who Had an Acute Symptom in the Past 12 Months

A total of 130 participants (87/130, 45.3%) had an acute symptom in the past 12 months. Their first action was mostly consulting with a physician (52/130, 40.0%) followed by self-medication (45/130, 34.6%), searching the internet for health information (25/130, 19.2%), and seeking a pharmacist's advice (8/130, 6.2%). Two-thirds of them (85/130, 65.3%) improved in the first 3-4 days after the onset of an acute symptom irrespective of their first action. The most common symptoms were headache, back pain, and diarrhea. Participants who had frequent doctor's visits, a doctor with whom they consult regularly, or a chronic medical illness were more likely to first seek either Web-based health information or a doctor's help rather than self-medicate or seek a pharmacist's advice (Table 3).

Among participants who searched the internet first (n=25), almost half (9/18) consulted a doctor because of the obtained Web-based information; the other half either self-medicated or sought a pharmacist's advice. Most of them (18/24, 75%) improved in the first 3-4 days.

Among those who self-medicated after searching the internet (n=6), only one participant's condition did not improve. Similarly, among those who sought care from a doctor after searching the internet (n=8), two participants' condition did not improve and one participant's condition worsened. Although many participants (22/25, 88%) found the Web-based information helpful to understand their acute symptoms, two-thirds (15/25, 60%) became anxious, and approximately 72% (18/25) discussed the obtained information with their doctor.

Finally, data analyses were conducted to compare the various demographics and actions of participants among those recruited through Google Ads versus through phone-based interviews. It was shown that participants were similar in both groups, with the exception that those recruited via Google Ads reported higher trust in Web-based health information and were more educated than those recruited via phone-based interviews.

Table 3. Bivariate analyses of the first action of the participants with an acute symptom and the various variables.

Variables	First action after an acute symptom				*P* value[a]
	Sought Web-based health information	Sought a doctor's care	Self-medicated	Sought a pharmacist's advice	
Doctors' visits in the past 12 months (N=121), n (%)					.002
0	1 (4.8)	1 (2.0)	11 (26.2)	2 (25.0)	
1-2	8 (38.1)	15 (30.0)	21 (50.0)	4 (50.0)	
3	3 (14.3)	10 (20.0)	3 (7.1)	0 (0.0)	
≥4	9 (42.9)	24 (48.0)	7 (16.7)	2 (25.0)	
Presence of a chronic medical condition (N=121), n (%)					.05
Yes	10 (47.6)	23 (46.0)	10 (23.8)	1 (12.5)	
No	11 (52.4)	27 (54.0)	32 (76.2)	7 (87.5)	
Presence of a physician who they consult on a regular basis (N=122), n (%)					.008
Yes	16 (76.2)	37 (72.5)	24 (55.8)	1 (14.3)	
No	5 (23.8)	14 (27.5)	19 (44.2)	6 (85.7)	
Status of perceived health, n (%)					.20
Very good	5 (23.8)	8 (16.0)	11 (25.6)	2 (25.0)	
Good	10 (47.6)	17 (34.0)	16 (37.2)	3 (37.5)	
Moderate	5 (23.8)	17 (34.0)	14 (32.6)	0 (0)	
Poor	1 (4.8)	7 (14.0)	1 (2.3)	3 (37.5)	
Very poor	0 (0)	1 (2.0)	1 (2.3)	0 (0)	

[a]Chi-square test.

Discussion

Principal Findings

Web-based health information seeking is well studied in the context of chronic medical conditions; however, only a few studies analyzed Web-based health information seekers for an acute symptom and the impact of such information on one's health. This study aimed to add to the literature by examining the prevalence of Web-based health information seeking in the context of an acute illness through a survey. This survey found that many participants (173/277, 62.5%) used the internet for an acute health symptom. Specifically, almost one-fifth (25/130) of those who had an acute symptom in the past 12 months sought Web-based health information first; among those people, two-thirds (9/25, 36.0%) eventually sought care from a doctor based on the information obtained. The majority improved, and the few of them who did not improve or worsened were among those who visited the doctor. The decision to first check Web-based health information was more prevalent among participants who had a doctor whom they consulted regularly, had frequent doctor's visits in the past 12 months, and had a chronic medical condition.

This study has shown that almost two-thirds (173/277, 62.5%) of the participants used the internet for an acute symptom. This is consistent with the rates of internet usage for general health-related information [3,12,24-26]. Moreover, in comparison with prior studies, Web-based health information seekers are younger [12,25,27,28], highly educated [25,27,28], and consider themselves healthy [27,28]. However, in this study, those who searched the internet for acute symptoms were found to have a middle-level income and not a high income, as seen for general Web-based health information seeking in previous studies [3,25,27,28]. This could be explained by the fact that many participants used Web-based information to self-manage their symptoms. Self-diagnosis and self-medication have been well associated with the socioeconomic status of patients [29-31]. Seeking health information for acute symptoms may

have different motives than seeking general health information for chronic or established diseases.

A good number of participants in this survey (25/130, 19.2%) sought Web-based health information first when they had an acute symptom, which is similar to what was found in a study that analyzed the search queries to a patient education orthopedic website: 17% of the patients visited the website for symptoms as compared with 32% for a condition and 22% for a certain treatment [7]. Similarly, 31% of college students sought to use the internet for self-diagnosis [13]. The harms of self-diagnosis through Web-based health information are not well studied. A study among patients with colorectal cancer reported that 25% of the sample consulted the internet before visiting health care providers [14]. Among those who checked information on the internet, 8.2% (n=5) reported that the information influenced their thinking, as it indicated that their condition was not cancer or did not require medical attention. Most of the participants in this study improved after looking up health information on the internet related to their acute symptoms; this might be related to the natural course of most acute symptoms. However, those who did not improve or worsened were among those who visited the doctor after they obtained the information. It is possible that the information obtained was enough for them to decide if they can wait for a natural resolution of the symptoms or there is a need to visit a doctor. In general, the majority of Web-based health information seekers report that this kind of information was helpful and improved their health status or their medical information [4,12,32].

The bivariate analysis found that the decision to search the internet or visit a doctor first was associated with having a doctor with whom they consulted regularly and with a high frequency of doctor's visits. It seems that participants who did not establish care within the medical system are more likely to self-medicate or visit the pharmacist. Despite that, some participants would seek information on the internet first, and almost half of them who did ended up visiting the doctor. Although physicians remain the most trusted source of medical information [25,33], the literature is mixed about the effect of Web-based information on patients' confidence in their physicians [34-36]. Discordance between the information obtained from the internet and that from the physician may compromise the trust in the physician [37]. Nevertheless, physicians should encourage their patients to discuss health information obtained from the internet and should guide them toward trusted online websites [38] to help them play an active role in symptom or disease management and control their anxiety.

Limitations

The generalizability of the data is limited, as the study was conducted among Lebanese adults and had a low response rate. However, the data are a revelation of the magnitude of participants seeking Web-based health information for acute symptoms. Furthermore, two different recruitment methodologies were used sequentially, which lead to more educated participants reporting higher trust in Web-based health information in the Google Ads than in the phone-based interviews. Moreover, there was probably some recall bias, mainly in answering the question addressing health information seeking for the last acute event during the past 12 months.

This survey showed that Web-based health information seeking for acute symptoms is common. However, physicians remain an important source of trustable medical information. Despite seeking Web-based health information first, almost half of the participants eventually visited a doctor and discussed the information that they sought online.

Acknowledgments

We acknowledge Hiba Yassine and Sara Iskandarani (students) for their contribution in data collection as part of the Medical Research Volunteer Program. LA was affiliated with the American University of Beirut at the time of the trial and is currently affiliated with Beirut Arab University.

Conflicts of Interest

None declared.

References

1. Internet World Stats. 2017. World internet users stats URL: https://www.internetworldstats.com/stats.htm [accessed 2019-10-27]
2. Internet World Stats. 2016. Lebanon Internet Stats URL: http://www.internetworldstats.com/me/lb.htm [accessed 2019-10-29]
3. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
4. Reinfeld-Kirkman N, Kalucy E, Roeger L. The relationship between self-reported health status and the increasing likelihood of South Australians seeking Internet health information. Aust N Z J Public Health 2010 Aug;34(4):422-426. [doi: 10.1111/j.1753-6405.2010.00576.x] [Medline: 20649784]
5. Rice RE. Influences, usage, and outcomes of internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]

6. Dickerson S, Reinhart AM, Feeley TH, Bidani R, Rich E, Garg VK, et al. Patient internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 2004;11(6):499-504 [FREE Full text] [doi: 10.1197/jamia.M1460] [Medline: 15298993]
7. Shuyler KS, Knight KM. What are patients seeking when they turn to the Internet? Qualitative content analysis of questions asked by visitors to an orthopaedics Web site. J Med Internet Res 2003 Oct 10;5(4):e24 [FREE Full text] [doi: 10.2196/jmir.5.4.e24] [Medline: 14713652]
8. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Mar 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
9. Shaw RJ, Johnson CM. Health information seeking and social media use on the internet among people with diabetes. Online J Public Health Inform 2011;3(1):pii: ojphi.v3i1.3561 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
10. Paolino L, Genser L, Fritsch S, De' Angelis N, Azoulay D, Lazzati A. The web-surfing bariatric patient: the role of the internet in the decision-making process. Obes Surg 2015 Apr;25(4):738-743. [doi: 10.1007/s11695-015-1578-x] [Medline: 25623374]
11. Shen MJ, Dyson RC, D'Agostino TA, Ostroff JS, Dickler MN, Heerdt AS, et al. Cancer-related internet information communication between oncologists and patients with breast cancer: a qualitative study. Psychooncology 2015 Nov;24(11):1439-1447 [FREE Full text] [doi: 10.1002/pon.3752] [Medline: 25631285]
12. Lorence DP, Park H, Fox S. Assessing health consumerism on the web: a demographic profile of information-seeking behaviors. J Med Syst 2006 Aug;30(4):251-258. [doi: 10.1007/s10916-005-9004-x] [Medline: 16978004]
13. Basch CH, MacLean SA, Romero R, Ethan D. Health information seeking behavior among college students. J Community Health 2018 Dec;43(6):1094-1099. [doi: 10.1007/s10900-018-0526-9] [Medline: 29779074]
14. Thomson MD, Siminoff LA, Longo DR. Internet use for prediagnosis symptom appraisal by colorectal cancer patients. Health Educ Behav 2012 Oct;39(5):583-588 [FREE Full text] [doi: 10.1177/1090198111423941] [Medline: 21990571]
15. Bennadi D. Self-medication: A current challenge. J Basic Clin Pharm 2013 Dec;5(1):19-23 [FREE Full text] [doi: 10.4103/0976-0105.128253] [Medline: 24808684]
16. Ruiz ME. Risks of self-medication practices. Curr Drug Saf 2010 Oct;5(4):315-323. [doi: 10.2174/157488610792245966] [Medline: 20615179]
17. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the internet: reviewing a complex and evolving landscape. Australas Med J 2014;7(1):24-28 [FREE Full text] [doi: 10.4066/AMJ.2014.1900] [Medline: 24567763]
18. Jones RB, Goldsmith L, Williams CJ, Kamel Boulos MN. Accuracy of geographically targeted internet advertisements on Google AdWords for recruitment in a randomized trial. J Med Internet Res 2012 Jun 20;14(3):e84 [FREE Full text] [doi: 10.2196/jmir.1991] [Medline: 22718043]
19. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015 Jul 22;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
20. Gross MS, Liu NH, Contreras O, Muñoz RF, Leykin Y. Using Google AdWords for international multilingual recruitment to health research websites. J Med Internet Res 2014 Jan 20;16(1):e18 [FREE Full text] [doi: 10.2196/jmir.2986] [Medline: 24446166]
21. Arya R, Antonisamy B, Kumar S. Sample size estimation in prevalence studies. Indian J Pediatr 2012 Nov;79(11):1482-1488. [doi: 10.1007/s12098-012-0763-3] [Medline: 22562262]
22. Lwanga SK, Lemeshow S, World Health Organization. Sample Size Determination in Health Studies: A Practical Manual. Geneva: WHO; 1991.
23. Bacha NN, Bahous R. Foreign language education in Lebanon: a context of cultural and curricular complexities. J Lang Teach Res 2011;2(6):1320. [doi: 10.4304/jltr.2.6.1320-1328]
24. Powell JA, Darvell M, Gray JA. The doctor, the patient and the world-wide web: how the internet is changing healthcare. J R Soc Med 2003 Mar;96(2):74-76 [FREE Full text] [doi: 10.1258/jrsm.96.2.74] [Medline: 12562977]
25. 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]
26. Alkhatlan HM, Rahman KF, Aljazzaf BH. Factors affecting seeking health-related information through the internet among patients in Kuwait. Alexandria J Med 2018;54(4):331-336. [doi: 10.1016/j.ajme.2017.05.008]
27. Miller LM, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
28. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
29. Awad A, Eltayeb I, Matowe L, Thalib L. Self-medication with antibiotics and antimalarials in the community of Khartoum State, Sudan. J Pharm Pharm Sci 2005 Aug 12;8(2):326-331 [FREE Full text] [Medline: 16124943]
30. Torres N, Chibi B, Middleton L, Solomon V, Mashamba-Thompson T. Evidence of factors influencing self-medication with antibiotics in low and middle-income countries: a systematic scoping review. Public Health 2019 Mar;168:92-101. [doi: 10.1016/j.puhe.2018.11.018] [Medline: 30716570]

31. Alhomoud F, Aljamea Z, Almahasnah R, Alkhalifah K, Basalelah L, Alhomoud FK. Self-medication and self-prescription with antibiotics in the Middle East-do they really happen? A systematic review of the prevalence, possible reasons, and outcomes. Int J Infect Dis 2017 Apr;57:3-12 [FREE Full text] [doi: 10.1016/j.ijid.2017.01.014] [Medline: 28111172]
32. 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]
33. 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]
34. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
35. Tanis M, Hartmann T, Te Poel F. Online health anxiety and consultation satisfaction: a quantitative exploratory study on their relations. Patient Educ Couns 2016 Jul;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]
36. 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]
37. Lu T, Xu YC, Wallace S. Internet usage and patient's trust in physician during diagnoses: a knowledge power perspective. J Assoc Inf Sci Technol 2017 Sep 22;69(1):110-120. [doi: 10.1002/asi.23920]
38. Hou J, Shim M. The role of provider-patient communication and trust in online sources in Internet use for health-related activities. J Health Commun 2010;15(Suppl 3):186-199. [doi: 10.1080/10810730.2010.522691] [Medline: 21154093]

Edited by C Eichenberg; submitted 27.06.19; peer-reviewed by S Kim, J Farzi; comments to author 31.08.19; revised version received 10.09.19; accepted 14.10.19; published 10.01.20.

Please cite as:
Aoun L, Lakkis N, Antoun J
Prevalence and Outcomes of Web-Based Health Information Seeking for Acute Symptoms: Cross-Sectional Study
J Med Internet Res 2020;22(1):e15148
URL: https://www.jmir.org/2020/1/e15148
doi:10.2196/15148
PMID:31922490

Original Paper

Health Consumers' Daily Habit of Internet Banking Use as a Proxy for Understanding Health Information Sharing Behavior: Quasi-Experimental Approach

Hyeyoung Hah[1], PhD

Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

Corresponding Author:
Hyeyoung Hah, PhD
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3053484342
Email: hhah@fiu.edu

Abstract

Background: As the US health care system is embracing data-driven care, personal health information (PHI) has become a valuable resource for various health care stakeholders. In particularly, health consumers are expected to autonomously manage and share PHI with their health care partners. To date, there have been mixed views on the factors influencing individuals' health data–sharing behaviors.

Objective: This study aimed to identify a key factor to better understand health information sharing behavior from a health consumer's perspective. We focused on daily settings, wherein health data–sharing behavior becomes a part of individuals' daily information management activities. Considering the similarity between health and finance information management, we explicitly examined whether health consumers' daily habit of similar data sharing from the financial domain affects their PHI-sharing behaviors in various scenarios.

Methods: A Web-based survey was administered to US health consumers who have access to and experience in using the internet. We collected individual health consumers' intention to share PHI under varying contexts, habit of financial information management (operationalized as internet banking [IB] use in this paper), and the demographic information from the cross-sectional Web-based survey. To isolate the effect of daily IB on PHI-sharing behaviors in everyday contexts, propensity score matching was used to estimate the average treatment effect (ATE) and average treatment effect on the treated (ATET) regarding IB use. We balanced the treatment and control groups using caliper matching based on the observed confounding variables (ie, gender, income, health status, and access to primary care provider), all of which resulted in a minimal level of bias between unmatched and matched samples (bias <5%).

Results: A total of 339 responses were obtained from a cross-sectional Web-based survey. The ATET results showed that in terms of sharing contents, those who used IB daily were more likely to share general information (P=.01), current information (P=.003), and entire data (P=.04). Regarding occasions for sharing occasions, IB users were prone to share their information in all cases (P=.02). With regard to sharing recipients, daily IB users were more willing to share their personal health data with stakeholders who were not directly involved in their care, such as health administrators (P=.05). These results were qualitatively similar to the ATE results.

Conclusions: This study examined whether daily management of similar information (ie, personal financial information) changes health consumers' PHI-sharing behavior under varying sharing conditions. We demonstrated that daily financial information management can encourage health information sharing to a much broader extent, in several instances, and with many stakeholders. We call for more attention to this unobserved daily habit driven by the use of various nonhealth technologies, all of which can implicitly affect patterns and the extent of individuals' PHI-sharing behaviors.

(J Med Internet Res 2020;22(1):e15585) doi:10.2196/15585

KEYWORDS

habits; personal health information; information sharing; propensity score; average treatment effect; internet banking; observational data; quasi-experimental design

Introduction

As the US health care system is embracing data-driven care, personal health information (PHI) needs to be shared and managed across clinical settings by health consumers [1]. Under this circumstance, PHI, defined as "individually identifiable information relating to the past, present, or future health status of an individual [2]," has become a valuable resource for various health care stakeholders. Recently, the Office of the National Coordinator for Health Information Technology and the Centers for Medicare and Medicaid Services have focused on allowing consumers free and easy access to their health data and enabling them to share PHI even with large technology companies or their selective care counterparts [3]. As health consumers have more choice in care and treatment and electronic access to their structured and unstructured health information [4,5], health consumers are expected to decide the PHI that needs to be shared with their chosen partners and the manner in which it should be shared [6,7]. In other words, individual health consumers should be capable of managing and sharing their PHI across various care contexts.

Health consumers manage not only health information but also nonhealth information on a daily basis. As individuals perform day-to-day functioning in the areas of finance, communication, transportation, socialization, and entertainment [8,9], such activities generate data and require decision-making on accessing, storing, and sharing these data daily. Among these areas, financial information is known to be similar to health information because it is private, personal information that is needed to be shared with strangers (eg, financial advisor), as mandated by government and local policies, and is influenced by individual consumers' knowledge [10,11], along with its industry-wide movement in consumer-centered services that enables personal financial data to be shared across incumbent banks and nonbanks (ie, Fintech firms) [12]. Thus, individual financial consumers have more power and control over their own data and are expected to make informed decisions. Given that individual consumers can be both health and financial consumers simultaneously, the same individuals' financial behavior can be a proxy for understanding health information sharing, all of which can provide a useful vantage point for understanding health information–sharing behavior.

Against this backdrop, little attention has been paid to explore health consumers' health data sharing in the context of daily living. We propose that health consumers are likely to be attracted from similar experiences and may have formed a habit with repeated exposure to the similar tasks while executing health and nonhealth tasks [13]. Taken together, the aim of this paper was to examine health consumers' existing habits regarding financial information management of the willingness to share health data in various scenarios. More specifically, we focus on three characteristics of health information–sharing behavior with respect to the extent of sharing contents, constituents, and instances [14,15]. By explicitly exploring cross-domain habitual activities in data management, this paper contributes to the ongoing discussion of the expected role of health consumers under the data-driven care system.

Methods

Survey Procedure

We recruited participants who were or had the potential to manage and share their own health data electronically. As we particularly focused on individuals who can manage health and financial tasks by the use of relevant technologies, we needed a study sample in which we can capture their daily activities of various data management beyond health care settings. To this end, we contracted with a market research company that had access to the paid panel of health consumers across 50 states in the United States for administering a cross-sectional Web-based survey in 2017. Each participant was incentivized by the completion and quality of their response, which was mainly managed by the market research firm. Our Web-based survey incorporated multiple items that measure health information–sharing behaviors, including health consumers' intention to share their health information, information sharing contexts, demographic information, and daily technology use such as internet banking (IB) use. An institutional review board approval was obtained before survey distribution. As a result, a total of 339 responses were used for further analysis.

Survey Instruments

All survey items were sourced from the existing literature, as presented in Table 1. First, for capturing health information–sharing intention with respect to *sharing contents*, *sharing situations*, and *sharing constituents*, we adopted our survey items from the study by Whiddett et al [14] and Anderson and Agarwal [15]. More specifically, sharing contents contained 4 items about the types of information likely to be shared, ranging from general information to all information (including sensitive disease information). *Sharing situations* included four items about the instances in which health consumers are willing to transfer their health information, such as all instances or selective or emergency cases. Finally, *sharing constituents* involved 10 items about whom an individual is willing to share their information with such as physicians, health administrators, and insurance payers. To determine dimensionality of the items for the sharing constituents, we ran exploratory factor analysis (EFA) with varimax rotation using SPSS (IBM Corporation, Armonk, New York). Our EFA identified two factors in *sharing constituents* based on their direct involvement in the care: One dimension is general constituents who do not directly engage in the care (eg, community physicians, government, and health insurance companies), whereas another dimension captures care-related counterparts (eg, physicians, nurses, and pharmacists). The items for health data sharing were anchored on a 5-point Likert scale (1: *strongly disagree* and 5: *strongly agree*). Second, we measured use frequency of daily technology such as computer, internet, email, and IB (ie, "How frequently

do you use internet banking?"). These items were anchored on four scales (ie, daily, weekly, monthly, and never), whose anchor was adapted from a mobile banking survey administered by the International Finance Corporation.

Finally, demographic information was captured for gender, marital status, income, education, occupation, race, ethnicity, health condition (chronic disease), and having a primary doctor within the domicile [16].

Table 1. Survey items.

Types	Survey items[a, b]	Reference
Sharing contents	• General information • Current health information • Past health information • All health information	[14,15]
Sharing instances	• In all cases and instances • For the purposes of care delivery within the clinical setting • For the purposes of other than provision of care (eg, research or marketing) • In case of medical emergency conditions	[14,15]
Sharing constituents		
General constituents	• Other physicians (who are not involved in your care) at hospitals • Other community physicians not involved in your care • Health administrators (eg, managers), government agencies • Health care researchers • Health insurance companies	[14,15]
Care-related constituents	• Physicians (who are involved in your care) at hospitals • Other community physicians involved in your care (treating physicians) • Nurses • Pharmacists	[14,15]
Habitual use of internet banking	• Frequency of internet banking use[c]	[17]

[a]We adopted all items from the study by Whiddett et al [14,15].

[b]Information-sharing items are measured on a 5-point Likert scale anchoring on 1 (strongly disagree) to 5 (strongly agree).

[c]Frequency of daily technology use is measured on daily, weekly, and monthly scales adopted from the survey of International Finance Corporation [17].

Statistical Analysis

The objective of the study was to examine whether frequent use of IB affects when, what, and with whom health consumers are willing to share their own personal health data. However, in observational studies similar to this study, it is often a challenge to isolate the treatment effect because confounding factors can influence both treatment and outcome [18]. To rule out such a confounding effect, prior studies have used various statistical estimation methods such as regression and panel methods, matching estimators, instrumental variables, and regression discontinuity designs [19-21]. Among these methods, we were particularly interested in propensity score matching (PSM), as it has been widely used when randomized clinical trials are infeasible in health care research [20,21]. PSM matches the observed and possible outcomes per each object based on a propensity score—a conditional probability that each observation receives treatment based on a set of observed covariates [22,23]. After matching, average treatment effects (ATEs) can be calculated by averaging out such a difference between the observed and potential outcomes [24]. This method depends on balancing observable covariates in treatment and control groups to isolate and estimate the treatment effect in the presence of confounding effects [25].

In our data, health consumers' use of IB and health data sharing can be confounded by known factors, that is, demographic characteristics and health status. Following a step-by-step suggestion from the study by Becker and Ichino [26], we conducted PSM (refer to the study by Becker and Ichino [26] for a more detailed explanation), including choice of confounding variables, balancing propensity score and covariates between treatment and control groups, and calculating ATE within the evaluating criteria of bias (the difference between estimated treatment and true effect) and precision of estimated treatment effects. As a first step, it is necessary to choose the correct sets of variables that affect both treatment and outcome variables to better isolate treatment effects. Next, based on the choice of confounding variables, one needs to validate whether there is an overlap in propensity scores between treatment and control groups, which is necessary for drawing an inference by comparing these two groups. After balancing propensity scores in the groups, one needs to check the distribution of covariates within the blocks of propensity score to determine if the treatment and control groups within the block have a similar covariate distribution except for the variation of treatment variable. On the basis of this, the final step is to estimate ATE of focus—either ATE among paired samples within blocks of a propensity score or average treatment effect on the treated

(ATET) for the treated observations only [27]. Although there is no clear guideline for calculating sample size for PSM, matching one or two untreated subjects to each treated subject is recommended when using PSM [28].

Results

Characteristics of Survey Participants

As shown in Table 2, most of the respondents were working professionals (236/339, 69.6%), female (225/339, 66.4%), and white (269/339, 79.4%). A total of 33.3% (113/339) have chronic conditions, and interestingly, almost all participants believe they are computer literate (337/339, 99.4%). Table 2 presents the demographic characteristics of our survey respondents across three user groups—IB users with daily, weekly, or monthly use frequencies. A majority of respondents use IB weekly (170/339, 50.1%), and participants aged 25-44 years were active users of IB across three user groups (approximately 203/339, 59.9%). We did not find any statistical group difference among demographic characteristics by *t* test ($P<.05$).

Although our data were obtained from a cross-sectional Web-based survey from US health consumers, we further evaluated the representativeness of our sample compared with established benchmark. The Board of Governors of the Federal Reserve System has conducted a Web-based survey on financial consumers' use of mobile banking in selective years [29]. As we examine the effect of IB use on PHI-sharing behavior, we compared the national profiles of IB consumers from the Federal Reserve Board in 2015 with our sample. We found that age distribution was particularly similar to our sample, as a majority of respondents were aged 25-65 years and a majority of users of mobile banking were aged 25-44 years (Multimedia Appendix 1). Furthermore, we checked age group differences among health consumers with IB frequency using a *t* test and found no significant group differences ($P<.05$).

Table 2. Characteristics of survey participants (total number of responses=339).

Demographic variables	All IB[a] users (N=339), n (%)	Daily IB users (n=96), n (%)	Weekly IB users (n=170), n (%)	Monthly IB users (n=73), n (%)
Gender[b]				
Male	114 (33.6)	34 (35)	54 (31.8)	26 (36)
Female	225 (66.4)	62 (65)	116 (68.2)	47 (64)
Marital status				
Married	188 (55.5)	51 (53)	103 (60.6)	34 (47)
Divorced	26 (7.7)	6 (6)	12 (7.1)	8 (11)
Separated	7 (2.1)	0 (0)	4 (2.4)	3 (4)
Never married	118 (34.8)	39 (41)	51 (30.0)	28 (39)
Age (years)				
18-24	53 (15.6)	15 (16)	21 (12.4)	17 (23)
25-34	128 (37.8)	39 (41)	67 (39.4)	22 (30)
35-44	78 (23.0)	23 (24)	42 (24.7)	13 (18)
45-54	43 (12.7)	13 (14)	24 (14.1)	6 (8)
55-64	27 (8.0)	5 (5)	11 (6.5)	11 (15)
≥65	10 (3.0)	1 (1)	5 (3.0)	4 (6)
Income status (US $)				
<20,000	51 (15.0)	13 (14)	23 (13.5)	15 (21)
20,000-39,999	76 (22.4)	20 (21)	30 (17.7)	26 (36)
40,000-59,999	59 (17.4)	22 (23)	28 (16.5)	9 (12)
60,000-79,999	53 (15.6)	14 (15)	31 (18.2)	8 (11)
80,000-99,999	44 (13.0)	7 (7)	31 (18.2)	6 (8)
>100,000	56 (16.5)	20 (21)	27 (15.9)	9 (12)
Education				
Less than high school	9 (2.7)	4 (4)	4 (2.4)	1 (1)
High school graduate	70 (20.7)	19 (20)	29 (17.1)	22 (30)
Some college	93 (27.4)	28 (29)	46 (27.1)	19 (26)
2-year degree	35 (10.3)	11 (12)	14 (8.2)	10 (14)
4-year degree	85 (25.1)	21 (22)	54 (31.8)	10 (14)
Master’s degree	40 (11.8)	10 (10)	21 (12.4)	9 (12)
PhD	7 (2.1)	3 (3)	2 (1.2)	2 (3)
Occupation				
Employed full time	196 (57.8)	65 (68)	102 (60.0)	29 (40)
Employed part time	40 (1.8)	9 (9)	21 (12.4)	10 (14)
Unemployed looking for work	29 (8.6)	7 (7)	11 (6.5)	11 (15)
Unemployed not looking for work	34 (10.0)	10 (10)	15 (8.8)	9 (12)
Retired	19 (5.6)	1 (1)	9 (5.3)	9 (12)
Disabled	21 (6.2)	4 (4)	12 (7.1)	5 (7)
Race				
White	269 (79.4)	77 (80)	137 (80.6)	55 (75)
Black	35 (10.3)	6 (6)	15 (8.8)	14 (19)

Demographic variables	All IB[a] users (N=339), n (%)	Daily IB users (n=96), n (%)	Weekly IB users (n=170), n (%)	Monthly IB users (n=73), n (%)
Asian	23 (6.8)	9 (9)	12 (7.1)	2 (3)
Other	12 (3.5)	4 (4)	6 (3.5)	2 (3)
Ethnicity				
Hispanic	38 (11.2)	10 (10)	21 (12.4)	7 (10)
Non-Hispanic	301 (88.8)	86 (90)	149 (87.7)	66 (90)
Chronic conditions				
Yes	113 (33.3)	24 (25)	60 (35.3)	29 (40)
No	226 (66.7)	72 (75)	110 (64.7)	44 (60)
Primary care access (distance)				
Within 5 miles	150 (44.3)	44 (46)	77 (45.3)	29 (40)
Within 10 miles	129 (38.1)	32 (33)	66 (38.8)	31 (43)
Within 30 miles	44 (13.0)	15 (16)	20 (11.8)	9 (12)
Not available	16 (4.7)	5 (5)	7 (4.2)	4 (6)

[a]IB: internet banking.

[b]n=236 for all IB users.

Propensity Score Matching Results

We conducted PSM analysis using Stata version 14.2 software (College Station, Texas). In our analysis, we chose income, race, health status, and having a primary care doctor within close proximity as our confounding variables, among others, which are likely to influence both IB use and health information sharing behavior. Subsequently, the propensity score was calculated for each block using a logit model. Figure 1 presents the distribution of propensity scores between the treated and untreated groups within blocks, termed as common support (propensity score overlaps in the matched pairs for each block), with appropriate overlap across blocks of propensity scores [30].

Table 3 illustrates covariate balancing among the variables before and after matching. It indicates that standard bias (percentage of bias) was less than 5% after caliper matching, and it is reasonable to move forward with the next step of analysis. To match a treated individual with nontreated individuals using similar propensity score, we used caliper matching (0.2×standard deviation of logit of propensity score with 1:2 neighbor matching with replacement) [26]. Figure 2 demonstrates a similar distribution of information-sharing behavior (sharing content) at baseline before and after caliper matching. Thus, we proceeded with ATE and ATET.

Figure 1. Distribution of propensity score between treated and untreated groups.

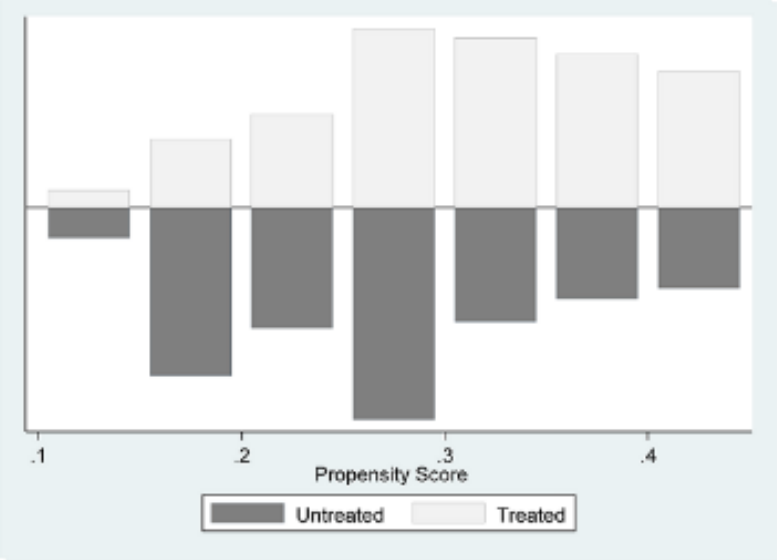
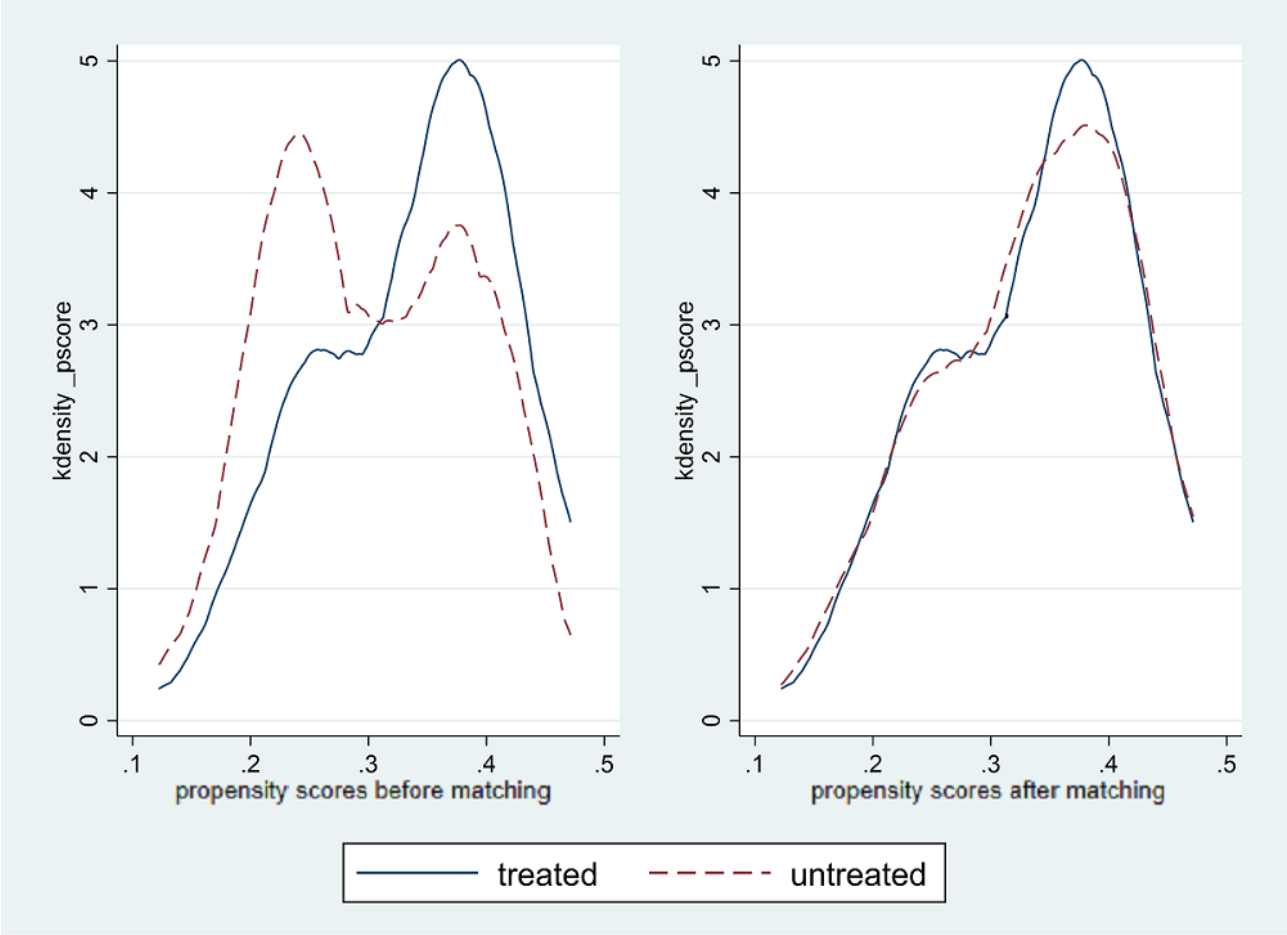


Table 3. Covariate balance before and after propensity score matching.

Variables	Unmatched sample				Matched sample			
	Mean, Treated	Mean, Untreated	Bias (%)	*P* value	Mean, Treated	Mean, Untreated	Bias (%)	*P* value
Gender								
Female	0.65	0.67	−5	.69	0.65	0.65	0	>.99
Income								
20,000-39,999	0.22	0.23	−1.6	.90	0.22	0.22	0	>.99
40,000-59,999	0.21	0.15	15.4	.21	0.21	0.21	0	>.99
60,000-79,999	0.16	0.16	1.1	.93	0.16	0.16	0	>.99
80,000-99,999	0.08	0.15	−21.8	.10	0.08	0.08	0	>.99
>100,000	0.22	0.15	19.4	.11	0.22	0.22	0	>.99
Chronic condition								
No	0.72	0.63	17.9	.16	0.72	0.72	0	>.99
Primary care access (distance)								
Within 5 miles	0.42	0.44	−2.6	.84	0.42	0.42	0	>.99

Figure 2. Density plots in health information sharing behavior. Treated sample comprised daily users of internet banking, and the rest of the users were included in the untreated group.



Average Treatment Effects of Daily Internet Banking Use

Finally, we estimated ATE and ATET as displayed in Tables 4 and 5, respectively. Here, ATE describes the ATE of IB use on health information sharing by comparing the treated (daily use of IB) and untreated groups (nondaily use of IB) in the matched pair sample. ATET indicates the estimated average effect of daily use of IB on health information sharing among daily IB users. We defined a treatment effect of IB as the use of IB on a daily basis, as we viewed daily use of IB as health consumers' daily habit. Prior literature noted that habit is an automatic

response to be formed by more frequent use and exposure of a focal event, technologies, and tasks [31]. Thus, we can consider daily use of IB as an existing habit of managing private personal information compared with weekly or monthly use of IB in our data. As such, we proceed to examine the treatment effect of daily use of IB on health information sharing under various circumstances.

In Table 4, there are differential behaviors in sharing contents, situations, and constituents between the treated and untreated groups. Daily IB users were more willing to share general and current health information (P=.009 and P=.004, respectively) for all situations (P=.06). For sharing constituents, these individuals were prone to share their information with non–care-related personnel, such as health administrators at a hospital (P=.05). Table 5 presents the average effect of daily IB use on health information sharing among daily IB users. We found that the treated individuals were likely to share all PHI with sensitive contents (P=.04) for all situations (P=.02). Moreover, they shared more PHI with non–care-related health administrators within the hospital (P=.05). Overall, results from ATE and ATET were qualitatively similar.

Table 4. Average treatment effect of daily internet banking use for matched pair sample.

Outcomes	Coefficient	SE	Z score	*P* value	95% CI
Sharing contents					
General information	0.324	0.124	2.61	.009	0.081 to 0.567
Current information	0.364	0.125	2.91	.004	0.119 to 0.61
Past information	0.17	0.127	1.34	.18	−0.08 to 0.42
Full information	0.215	0.153	1.4	.16	−0.085 to 0.514
Sharing instances					
All cases and situations	0.281	0.151	1.87	.06	−0.014 to 0.577
Care purposes	0.131	0.111	1.18	.24	−0.086 to 0.349
Noncare purposes	0.086	0.178	0.48	.63	−0.263 to 0.435
Medical emergency	0.133	0.095	1.39	.16	−0.054 to 0.32
Sharing constituents					
For direct care					
Your physician	0.084	0.114	0.74	.46	−0.139 to 0.307
Involving community physician	−0.064	0.122	−0.53	.60	−0.304 to 0.175
Nurses	0.119	0.139	0.85	.39	−0.154 to 0.393
Pharmacists	−0.023	0.174	−0.13	.89	−0.364 to 0.317
For indirect care					
Noninvolving physician at hospital	0.278	0.191	1.45	.15	−0.097 to 0.653
Noninvolving community physician	0.234	0.187	1.25	.21	−0.132 to 0.601
Health administrators (eg, managers)	0.347	0.174	1.99	.05	0.005 to 0.688
Government	0.23	0.189	1.22	.22	−0.14 to 0.601
Health care researchers	0.179	0.185	0.96	.34	−0.185 to 0.542
Insurance	0.169	0.179	0.94	.35	−0.183 to 0.521

Table 5. Average treatment effects on the treated of daily internet banking use.

Outcomes	Coefficient	SE	Z score	*P* value	95% CI
Sharing contents					
General information	0.346	0.140	2.470	.01	0.071 to 0.621
Current information	0.399	0.134	2.960	.003	0.135 to 0.662
Past information	0.208	0.145	1.430	.15	−0.076 to 0.492
Full information	0.334	0.160	2.090	.04	0.021 to 0.647
Sharing instances					
All cases and situations	0.319	0.139	2.300	.02	0.047 to 0.591
Care purposes	0.192	0.117	1.640	.10	−0.037 to 0.421
Noncare purposes	0.156	0.200	0.780	.44	−0.236 to 0.547
Medical emergency	0.179	0.104	1.710	.09	−0.026 to 0.383
Sharing constituents					
For direct care					
Your physician	0.058	0.136	0.430	.67	−0.208 to 0.324
Involving community physician	−0.119	0.139	−0.860	.39	−0.392 to 0.153
Nurses	0.184	0.139	1.320	.19	−0.089 to 0.457
Pharmacists	0.021	0.155	0.140	.89	−0.283 to 0.326
For indirect care					
Noninvolving physician at hospital	0.331	0.200	1.660	.10	−0.060 to 0.722
Noninvolving community physician	0.201	0.199	1.010	.31	−0.188 to 0.590
Health administrators (eg, managers)	0.350	0.177	1.980	.05	0.003 to 0.698
Government	0.232	0.168	1.380	.17	−0.097 to 0.561
Health care researchers	0.146	0.177	0.820	.41	−0.202 to 0.493
Insurance	0.249	0.152	1.640	.10	−0.049 to 0.548

Discussion

Principal Findings

Despite growing expectation of individuals' responsibility for sharing PHI for data-driven care, there have been mixed results regarding factors influencing health consumers' sharing intention under various circumstances. Given that management of financial information resembles that of health information, this study called for more attention on individuals' daily use of financial information management as a proxy for health information sharing and hypothesized that the more frequently individuals managed their financial information, the more likely they were willing to share PHI under various sharing conditions. Our PSM results revealed that daily IB users were more willing to share the large extent of health information for all instances, even with personnel who were not directly involved in their care process. This is one of the first studies that explores the role of a daily habit of financial information management to predict individuals' intention in sharing their health information.

Limitations

Although we presented important findings on the role of daily habit of IB use, our results should be interpreted with caution because of their limitations. First, our study is conducted via a cross-sectional Web-based survey. Although our research question was aligned with our study design, tracking health consumers' information management behavior over time can provide an in-depth view and further identify contextual factors in the daily living context. The cross-section time series information on health consumers' information management in daily living would be beneficial in the future to capture granular level of measures and to control unobserved heterogeneity among individuals. Second, we conceptualized our treatment effect as habitual use of financial data management and operationalized it by the daily use of IB. Although our unidimensional, binary measure of IB use was appropriate for PSM methodology, future research can incorporate multi-item measures to capture multidimensional aspects of financial data management for health consumers in the richer research models. Finally, we acknowledge that majority of survey respondents in our study have no immediate health issues (237/339, 69.9%); therefore, the results of this study may not be generalized to patients who have various health conditions and statuses. As we assumed that the same individuals can be both health and financial consumers, the findings of this study can be a baseline information to compare individuals' behaviors in medical situations in subsequent research.

Comparison With Prior Work

Role of the Existing Habit in Daily Living

In this paper, we first examined the daily habit that has not been of focus in health care research. More specifically, this paper juxtaposed the similarity between health and financial data and identified financial data habit as a key factor in understanding health data sharing from the same individuals. Our approach assumed that the same individuals are customers for both health and financial services; therefore, such individual-level behaviors can be closely related. Theoretically, it is known that when people cope with a new event, their reaction might be predictable simply because there are likely to base their reaction on past experience or their knowledge of similar situations [32,33]. Defined as an automatic reaction toward certain stimuli or inputs based on past experience or learning [34,35], habit has been a key research variable to predict a certain behavior in education, health care, and information systems disciplines [10,36]. In the information search context, people seek information from easily available internet sources or acquaintances (eg, friends and family), and they typically return to such habits for future information-seeking behavior [33]. In technology use context, prior learning and habits from technology can influence adoption and use of a new technology [37,38]. Another example noted that health consumers' health status and Web-based community membership (ie, PatientsLikeMe) makes them feel comfortable sharing their sensitive health information with people who are not directly related to their care [39]. Thus, an individual's habitual behavior in one life area can affect the same individual's behavior in another life setting.

Although prior health care literature highlighted the importance of individuals' habit to understand health behaviors, habit has been mainly defined within the context of health care. For example, health consumers' exposure to personal health records technology influenced individuals' share of PHI with care providers and non–care-related providers [35,40]. Given the complexity of PHI and heterogeneous information sharing behaviors, there is a growing interest in understanding health consumers' differential information sharing. Previous seminal papers have documented that sharing of the most sensitive PHI varies widely under various conditions, with sharing counterparts—care-involved personnel (eg, family and caregivers) or a broad audience (government or non–care-related providers) [41-43]. Although prior studies point out to their existing predisposition on data security [44], familiarity with technology use [35,43] or Web-based community membership [40] as plausible reasons for such heterogeneous responses, these factors have been searched and identified in clinical settings. By recontextualizing health data sharing into everyday living contexts, we explicitly examined whether decisions on health information sharing are influenced by daily, nonhealth-related habit of individual consumers.

Lesson Learned From the Similarity Between Health and Financial Data

This paper also showed that frequent exposure to IB is positively related to health data–sharing behavior. This finding is in line with and extends prior health research in two ways. First, although the effect of internet use has been widely discussed to understand health information sharing behavior, the influence of habitual internet use is lesser known [40]. At a granular level, we identified that daily users of IB are more prone to share more data for all cases and with non–care-related stakeholders. Second, this study identified opportunities to identify new phenomena from financial activities of individual health consumers. In the financial sector, individual consumers have more power and control over their own data and are expected to make informed decisions in various situations under the nationwide trends of consumer-driven service. A recent survey shows that 60% of consumers are willing to share personal data (eg, location data and lifestyle information) with financial service providers in exchange for customized promotion and better services, and young tech-savvy customers are willing to share more data [45]. Yet, such data sharing is based on each consumer's autonomous decision about sharing across traditional banks and nonbanks (ie, Fintechs) [46]. Viewing health information sharing as a part of multidimensional information managing tasks in everyday living, this study clearly demonstrated that a cross-domain habit of information management is positively associated with health information sharing, which has largely been underexplored in health care research [47].

Future Research

For future research, it will be worthwhile to revisit this research model in clinical setting and explore the effect of habit on health data sharing in the clinical setting for those who have chronic conditions or medical urgency. Health data sharing is a complex and variable phenomenon, and more interdisciplinary research is indispensable. As health consumers are attaining more ownership to manage and share their own health information via websites, wearables, and mobile apps, it is important to determine whether they are capable of dealing with this volume of data [48] and whether they make informed decisions to share personal data with multiple stakeholders. Various types of structured and unstructured health information may be stored at an individual's home or workplace or at a hospital. Given that such data need to be freed by the hands of health consumers and distributed for the optimal health care decision and outcomes [49], future research can further explore various information type and their willingness to share such data under different sharing scenarios.

Practical Implications

This research has practical implications. First, health technology vendors may design health information management tools modeled after financial information management tools [50]. As health care and finance are both characterized by the important role of system usability for customer satisfaction, involving more consumers in the rapid process of innovation and understanding consumer behaviors and fast-moving information technology trends [51], system design can support health consumers' willingness to share their own data and further liberate their data for consumer-centered care [52]. To better manage health information for self-health management [53], individual health consumers need to collectively as well as selectively manage either type of health information and use it for their own medical conditions and contexts. Health

technology can support such sharing behavior by providing repository of health data that is similar to a consolidated bank account and allowing them to selectively share or store it [51]. Second, health consumers need to be educated on how to manage various types of health information, including access, process, and exchange of public and private health information. As health information is domain-specific and either personal [15] or widely available on the internet [54], individual health consumers should understand the difference between public and private health information and shareability of such information accordingly. As shown in our results, even with the experience of managing sensitive financial data daily, health consumers may not be ready to share PHI with other stakeholders who are not directly involved in their care. This has timely implications on the direction and content of health consumer education about sharing constituents [55].

Conclusions

The objective of this study was to examine the effect of daily use of financial technology on health information sharing. Considering the similarity between health and financial technology and the characteristics of such information, this study proposes that the unobserved habit of managing sensitive information daily can further affect managing and sharing another type of sensitive information—PHI. Results from PSM reveal that frequent users of financial technology are more prone to share their entire health information in all instances, even with non–care-related stakeholders. Subsequent research can explore more granular types of habits in various life domains to better understand health consumers' readiness to manage self-health information for realizing consumer-centered care in the future.

Acknowledgments

We would like to thank Pouyan Zadeh for the support in data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
National profiles of mobile banking users.
[DOCX File , 13 KB - jmir_v22i1e15585_app1.docx]

References

1. Adler-Milstein J, Jha AK. Sharing clinical data electronically: a critical challenge for fixing the health care system. J Am Med Assoc 2012 Apr 25;307(16):1695-1696. [doi: 10.1001/jama.2012.525] [Medline: 22535851]
2. HIPAA Journal. 2018. What is Considered Protected Health Information Under HIPAA? URL: https://www.hipaajournal.com/what-is-considered-protected-health-information-under-hipaa/ [accessed 2019-10-02]
3. Center for Medicare and Medicaid Services (CMS). 2019 Feb 8. CMS Advances Interoperability & Patient Access to Health Data through New Proposals URL: https://www.cms.gov/newsroom/fact-sheets/cms-advances-interoperability-patient-access-health-data-through-new-proposals [accessed 2019-10-02]
4. HealthIT. 2019. Notice of Proposed Rulemaking to Improve the Interoperability of Health Information URL: https://www.healthit.gov/topic/laws-regulation-and-policy/notice-proposed-rulemaking-improve-interoperability-health [accessed 2019-10-02]
5. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
6. Blumenthal D. Harvard Business Review. 2019 Mar 15. A Big Step Toward Giving Patients Control Over Their Health Care Data URL: https://hbr.org/2019/03/a-big-step-toward-giving-patients-control-over-their-health-care-data [accessed 2019-10-02]
7. Wolff J, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 Apr 1;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
8. Melrose RJ, Brommelhoff JA, Narvaez T, Natta L, Riskin-Jones HH, Sakhai S, et al. The use of Information and Communication technology when completing instrumental activities of daily living. Comput Hum Behav 2016;63:471-479. [doi: 10.1016/j.chb.2016.05.045]
9. Backonja U, Kim K, Casper G, Patton T, Ramly E, Brennan P. Observations of daily living: putting the 'personal in personal health records. NI 2012 (2012) 2012;2012:6 [FREE Full text] [Medline: 24199037]
10. Eiser J, Morgan M, Gammage P, Brooks N, Kirby R. Adolescent health behaviour and similarity-attraction: friends share smoking habits (really), but much else besides. Br J Soc Psychol 1991 Dec;30 (Pt 4):339-348. [doi: 10.1111/j.2044-8309.1991.tb00950.x] [Medline: 1799862]

11. Fasciani L. Healthcare Global. 2018 Aug 5. From Finance to Healthcare: 5 Fintech Trends That Will Benefit Digital Health Products URL: https://www.healthcareglobal.com/technology/finance-healthcare-5-fintech-trends-will-benefit-digital-health-products [accessed 2019-10-02]
12. Health IT Analytics. 2018. Why Healthcare Should Study Up on Lessons Learned in Fintech URL: https://healthitanalytics.com/news/why-healthcare-should-study-up-on-lessons-learned-in-fintech [accessed 2019-10-02]
13. Continuum Innovation. 2015. Five Things that Healthcare Can Learn from Financial Services URL: https://www.continuuminnovation.com/en/how-we-think/blog/five-things-that-healthcare-can-learn-from-financial-services [accessed 2019-10-02]
14. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. Int J Med Inform 2006 Jul;75(7):530-541. [doi: 10.1016/j.ijmedinf.2005.08.009] [Medline: 16198142]
15. Anderson CL, Agarwal R. The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. Inf Syst Res 2011;22(3):469-490. [doi: 10.1287/isre.1100.0335]
16. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 2;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
17. McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 2004 Dec;9(4):403-425. [doi: 10.1037/1082-989X.9.4.403] [Medline: 15598095]
18. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. J Am Med Assoc 2007 Jan 17;297(3):278-285 [FREE Full text] [doi: 10.1001/jama.297.3.278] [Medline: 17227979]
19. Nichols A. Erratum and discussion of propensity-score reweighting. Stata J 2008;8(4):532-539. [doi: 10.1177/1536867x0800800405]
20. D'Agostino RB, Kwan H. Measuring effectiveness. What to expect without a randomized control group. Med Care 1995 Apr;33(4 Suppl):AS95-A105. [Medline: 7723466]
21. D'Agostino Jr RB, D'Agostino Sr RB. Estimating treatment effects using observational data. J Am Med Assoc 2007 Jan 17;297(3):314-316. [doi: 10.1001/jama.297.3.314] [Medline: 17227985]
22. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79(387):516-524. [doi: 10.1080/01621459.1984.10478078]
23. Abadie A, Imbens GW. Matching on the estimated propensity score. Econometrica 2016;84(2):781-807. [doi: 10.3982/ecta11293]
24. International Finance Corporation. Mobile Banking Questionnaire URL: https://www.ifc.org/wps/wcm/connect/502ff234-394c-4a04-b886-a8adbf8d1cf1/Tool+3.9.+Market+Questionnaire+-+Users.pdf?MOD=AJPERES&CVID=jMCOjO0 [accessed 2019-10-02]
25. Jalan J, Ravallion M. Estimating the benefit incidence of an antipoverty program by propensity-score matching. J Bus Econ Stat 2003;21(1):19-30. [doi: 10.1198/073500102288618720]
26. Garrido MM, Kelley AS, Paris J, Roza K, Meier DE, Morrison RS, et al. Methods for constructing and assessing propensity scores. Health Serv Res 2014 Oct;49(5):1701-1720 [FREE Full text] [doi: 10.1111/1475-6773.12182] [Medline: 24779867]
27. Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. Stata J 2002;2(4):358-377. [doi: 10.1177/1536867x0200200403]
28. Austin PC. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. Am J Epidemiol 2010 Nov 1;172(9):1092-1097 [FREE Full text] [doi: 10.1093/aje/kwq224] [Medline: 20802241]
29. Federal Reserve Systems. 2018 Mar 27. Mobile Banking: A Closer Look at Survey Measures URL: https://www.federalreserve.gov/econres/notes/feds-notes/mobile-banking-a-closer-look-at-survey-measures-20180327.htm [accessed 2019-10-02]
30. Leuven E, Sianesi B. IDEAS/RePEc: Economics and Finance Research. 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Version 3.0.0 URL: https://ideas.repec.org/c/boc/bocode/s432001.html [accessed 2019-10-02]
31. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. Manag Inf Syst Q 2012;36(1):157-178. [doi: 10.2307/41410412]
32. Dewdney P, Harris R. Barriers To Information: How Formal Help Systems Fail Battered Women. Westport, Conn: Praeger; 1994.
33. Fisher KE, Naumer CM. Information grounds: Theoretical basis and empirical findings on information flow in social settings. In: Amanda S, Charles C, editors. New Directions In Human Information Behavior. Dordrecht: Springer; 2006:93-111.
34. Limayem M, Hirt S. Force of habit and information systems usage: theory and initial validation. J Assoc Inf Syst 2003;4(3):65-97. [doi: 10.17705/1jais.00030]

35. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
36. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016;17(5):328-376. [doi: 10.17705/1jais.00428]
37. Wang C, Harris J, Patterson P. The roles of habit, self-efficacy, and satisfaction in driving continued use of self-service technologies. J Serv Res 2013;16(3):400-414. [doi: 10.1177/1094670512473200]
38. Limayem M, Cheung CM. Understanding information systems continuance: the case of internet-based learning technologies. Inf Manag 2008;45(4):227-232. [doi: 10.1016/j.im.2008.02.005]
39. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
40. Frost J, Vermeulen I, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. J Med Internet Res 2014 May 14;16(5):e126 [FREE Full text] [doi: 10.2196/jmir.2684] [Medline: 24828114]
41. Jackson SL, Mejilla R, Darer JD, Oster NV, Ralston JD, Leveille SG, et al. Patients who share transparent visit notes with others: characteristics, risks, and benefits. J Med Internet Res 2014 Nov 12;16(11):e247 [FREE Full text] [doi: 10.2196/jmir.3363] [Medline: 25405911]
42. Ivanov A, Sharman R, Rao HR. Exploring factors impacting sharing health-tracking records. Health Policy Technol 2015;4(3):263-276 [FREE Full text] [doi: 10.1016/j.hlpt.2015.04.008]
43. Weitzman ER, Kelemen S, Kaci L, Mandl KD. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak 2012 May 22;12:39 [FREE Full text] [doi: 10.1186/1472-6947-12-39] [Medline: 22616619]
44. Bansal G, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
45. Business Insider. 2019. Financial Service Consumers Are Willing to Share Their Personal Data for Benefits and Discounts URL: https://www.businessinsider.com/financial-service-consumers-share-personal-data-for-benefits-discounts-2019-3 [accessed 2019-10-02]
46. McKinsey & Company. 2017. Data sharing and open banking URL: https://www.mckinsey.it/sites/default/files/data-sharing-and-open-banking.pdf [accessed 2019-10-02]
47. Tamilmani K, Rana NP, Dwivedi YK. Use of 'habit' is not a habit in understanding individual technology adoption: a review of UTAUT2 based empirical studies. In: Elbanna A, Dwivedi YK, Bunker D, Wastell D, editors. Smart Working, Living and Organising. Cham: Springer; 2018:277-294.
48. Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. J Am Med Assoc 2015 Jan 27;313(4):355-356 [FREE Full text] [doi: 10.1001/jama.2014.10716] [Medline: 25408371]
49. Haug CJ. Whose data are they anyway? Can a patient perspective advance the data-sharing debate? N Engl J Med 2017 Jun 8;376(23):2203-2205. [doi: 10.1056/NEJMp1704485] [Medline: 28445103]
50. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
51. InformationWeek. 2014. 5 Lessons For Healthcare Providers From Banks URL: https://www.informationweek.com/healthcare/patient-tools/5-lessons-for-healthcare-providers-from-banks/d/d-id/1114185 [accessed 2019-10-02]
52. Academy Health. 2019. Health Data Liberator Award Honors Work to Help Consumers Make Smarter Choices URL: https://www.academyhealth.org/blog/2019-03/health-data-liberator-award-honors-work-help-consumers-make-smarter-choices [accessed 2019-10-02]
53. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
54. Eysenbach G. Consumer health informatics. Br Med J 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
55. Bracha Y, Bagwell J, Furberg R, Wald JS. Consumer-mediated data exchange for research: current state of US law, technology, and trust. JMIR Med Inform 2019 Jun 3;7(2):e12348 [FREE Full text] [doi: 10.2196/12348] [Medline: 30946692]

Abbreviations

ATE: average treatment effect
ATET: average treatment effect on the treated
EFA: exploratory factor analysis
IB: internet banking
PHI: personal health information
PSM: propensity score matching

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by J Papadakos, L Fleisher, A Samuel, M Rezai-Rad; comments to author 19.08.19; revised version received 04.10.19; accepted 31.10.19; published 08.01.20.

Please cite as:
Hah H
Health Consumers' Daily Habit of Internet Banking Use as a Proxy for Understanding Health Information Sharing Behavior: Quasi-Experimental Approach
J Med Internet Res 2020;22(1):e15585
URL: https://www.jmir.org/2020/1/e15585
doi:10.2196/15585
PMID:31913129

Original Paper

Effect of Patient-Physician Relationship on Withholding Information Behavior: Analysis of Health Information National Trends Survey (2011-2018) Data

Xin Yang[1], MSc, PhD; Jason Parton[2], PhD; Dwight Lewis[3], PhD; Ning Yang[2], MSc; Matthew Hudnall[2], PhD

[1]Institute of Business Analytics, The University of Alabama, Tuscaloosa, AL, United States

[2]Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL, United States

[3]Department of Management, The University of Alabama, Tuscaloosa, AL, United States

Corresponding Author:
Xin Yang, MSc, PhD
Institute of Business Analytics
The University of Alabama
254 Bidgood Hall
Tuscaloosa, AL, 35401
United States
Phone: 1 2053483267
Email: xyang15@cba.ua.edu

Abstract

Background: Patients' withholding information from doctors can undermine medical treatment, create barriers for appropriate diagnoses, and increase systemic cost in health care systems. To date, there is limited literature detailing the association between trends of patients withholding information behavior (WIB) and the patient-physician relationship (PPR).

Objective: The aim of this study was to explore the prevalence trend of WIB after 2011 and examine the effects of PPR on WIB and its time trend.

Methods: A total of 5 iterations of data from the Health Information National Trends Survey (years: 2011-2018; n=11,954) were used to explore curvilinear trends of WIB among the US population. Multiple logistic regression models were used to examine curvilinear time trends of WIB, effects of PPR on WIB, and moderation effects of PPR on the WIB time trend.

Results: The WIB prevalence has an increasing trend before 2014, which has the highest rate of 13.57%, and then it decreases after 2014 to 8.65%. The trend of WIB is curvilinear as the quadratic term in logistic regression model was statistically significant (P=.04; beta=−.022; SE=0.011; odds ratio [OR] 0.978, 95% CI 0.957-0.999). PPR is reversely associated with WIB (P<.001; beta=−.462; SE=0.097; OR 0.630, 95% CI 0.518-0.766) and has a significant moderation effect on time trends (P=.02; beta=−.06; SE=0.025; OR 0.941, 95% CI 0.896-0.989). In general, poor quality of PPR not only significantly increased the WIB probability but also postponed the change of point for WIB curvilinear trend.

Conclusions: Findings suggest that the time trend of WIB between 2011 and 2018 is curvilinear and moderated by the quality of the PPR. Given these results, providers may reduce WIB by improving PPR. More research is needed to confirm these findings.

(J Med Internet Res 2020;22(1):e16713) doi:10.2196/16713

KEYWORDS

withholding information behavior; patient-physician relationship; electronic medical records; privacy

Introduction

Background

Medical literature recognizes that quality communicative interactions between patients and their doctors are important to the success of medical treatment [1]. Many medical conditions can be accurately diagnosed by providers in an efficient manner if patients' medical history is fully divulged [2] and clinicians express some empathy during medical visits [3]. As such, limited disclosure of critical health information between patients and doctors serves as a barrier to the success of a prescribed medical regimen. Despite this intuitive knowledge, existing evidence suggests that patients purposefully withhold relevant medical information from clinicians [4].

The reasons for patients withholding information are multifold. Several studies have examined factors that influence withholding information behavior (WIB), including demographics [5], social economics [6], and health [7]. Additional reasons include embarrassment of maladaptive health behavior and aversion to sensitive medical topics that patients are reluctant to share comprehensive details about their health status [4]. Recent medical data breaches are another possible reason why patients may not share all information and materials related to their medical history. In particular, prior studies highlight that individuals are more likely to withhold information when there are increased concerns about security [8]. Between the years 2009 and 2018, there were 2546 reported health care data breaches, which affected the health care records of 189,945,874 patients [9]. Given the increase in prevalence of medical data breaches across the United States as well as the potential influence of patients' perceptions on medical data security via a willingness to share their medical history, an understanding of patients' withholding patterns is warranted. To date, the scholarly findings are limited on the topic of patients' WIB and its prevalence time trend in recent years. Evidence by Patel et al suggests that the proportion of US patients who exhibited medical WIB were 7%, 8%, and 5% for the years 2012, 2013, and 2014, respectively, indicating that the time trend (ie, slope) of WIB prevalence may not be linear or moving in a specific direction [10]. However, Patel et al's [10] findings appear limited because of the study's relatively small sample size and period, and the United States has witnessed a fair number of health care data breaches since publication. From our searches, we have found no additional studies that either support or conflict with these findings. Updated findings examining WIB that include more periods may provide better insights for policy-making state and national health officials.

Up to now, there is a paucity of literature on how patient-physician relationships (PPRs) influence patients' WIB. PPR, which is measured by time spent, understanding, involvement, and helpfulness [11,12], is thought to be positively related with the intention to share protected health information (PHI) [7]. In a qualitative study among a small group of female Latina patients, 26 out of 28 study participants reported that their willingness to disclose health information depends on a PPR including spending enough time with patients, involving patients in decisions, and paying attention to patients' feelings and emotions [13]. In a phenomenology (ie, the sciences of how people experience phenomena) study, factors such as being involved, listening attentively, and leaving time for patients influenced participants to disclose their use of traditional and complementary medicine to their doctors [14]. However, to the best of our knowledge, no study has empirically examined how PPRs influence patients' withholding behavior and its longitudinal trend in a nationwide representative sample of the US population. An examination of the effects of PPR on WIB time trends may provide valuable insights for policy makers as well as providers to inform them on potential needs to adjust communicative practices with prospective patients.

Objectives

Given the previously mentioned knowledge gaps about WIB in the United States, we investigate the longitudinal prevalence of WIB and the moderating effects of PPR on WIB trends adjusting for variables well studied in previous research. Using 2011 to 2018 responses from the Health Information National Trends Survey (HINTS), we sought to investigate the following areas:

> *Q1: Is there a stable pattern of WIB among patients over time?*
>
> *Q2: Does increased self-reported satisfaction of a patient's relationship with their provider meaningfully lower the odds that a patient would withhold important medical information during treatment?*
>
> *Q3: Does the PPR influence the national trajectory of WIB over time?*

We believe that findings from this study could contribute to shifting the paradigm in medical counseling research in several aspects. First, technological advancements and setbacks in the medical field have a dynamic and arguably instantaneous relationship with population health behavior [15]. Therefore, studies examining WIB and associated factors may be more informative if investigated over time. Moreover, increased knowledge of the influence that the PPR has on WIB may help health officials to allocate resources for behavioral and counseling-related patient-provider interventions to promote population health outcomes in the United States.

Methods

Sample

Findings from this study were derived using the following 5 waves of responses from HINTS: 2011(HINTS 4 cycle 1), 2012 (HINTS 4 cycle 2), 2014 (HINTS 4 cycle 4), 2017 (HINTS 5 cycle 1), and 2018 (HINTS 5 cycle 2). HINTS is an ongoing serial cross-sectional survey conducted by the National Cancer Institute to document attitudes and perceptions about health information access and use among noninstitutionalized US adults. The primary inclusion criteria for HINTS waves were based on the availability of survey questions focused on WIB. Participants self-reporting no visits to a nonemergency provider during the past 12 months were also excluded from analyses as these individuals would not have the opportunity to demonstrate WIB. A total sample of 11,954 respondents with complete records across 5 survey years, which represents about half US population each year, was used for descriptive analyses and logistic regression models.

Dependent Variable

Our outcome of interest (WIB) was obtained from the HINTS survey item: "Have you ever kept information from your health care provider because you were concerned about the privacy or security of your medical records? (Yes/No)" (Question D3, HINTS 5 cycle 2).

Patient-Physician Relationship

The PPR was operationalized using responses from the following 7 items in our study: (1) "How often did they give you the chance to ask all the health-related questions you had?," (2) "How often did they give the attention you needed to your feelings and emotions?," (3) "How often did they involve you in decisions about your health care as much as you wanted?,"

(4) “How often did they make sure you understood the things you needed to do to take care of your health?,” (5) “How often did they explain things in a way you could understand?,” (6) “How often did they spend enough time with you?,” and (7) “How often did they help you deal with feelings of uncertainty about your health or health care?.” Participant’s responses to PPR questions were measured using a Likert-type scale scored as follows: 1=always, 2=usually, 3=sometimes, and 4=never. In a previous study, these 7 items were used to define PPRs with a Cronbach alpha of .94 and composite reliability of 0.9 in confirmatory factor analysis [8]. In order to improve interpretability, we first reverse coded the 7 items (1=never, 2=sometimes, 3=usually, 4=always). To make logistic regression intercept meaningful, we substracted the reverse coded 7 items by 1 so that these 7 items had the minmum score of 0 (0=never, 1=sometimes, 2=usually, 3=always). A factor-based score reflecting PPR was generated by the mean of 7 recoded items’ scores and used as an independent variable in the logistic regression models [16]. To help readers internalize the context of our findings, the PPR was ordinally categorized into tertiles as “poor” (tertile 1: 0-2.14), “fair” (tertile 2: 2.15-2.85), and “good” (tertile 3: 2.86-3) in descriptive analyses. The PPR remained as a continuous variable in models adjusted for key covariates to optimize the statistical power in the study’s findings.

Other Variables

The time when surveys were conducted (2011, 2012, 2014, 2017, and 2018) was used to investigate the time trend aspect. Demographic variables include gender (male/female), age group (18-24, 25-44, 45-64, and 65+ years), census division (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific), born in the United States (yes/no), urbanity (urban and rural), occupation (employed, unemployed, retired, disabled, and other), race or ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, non-Hispanic other, and non-Hispanic Asian), and education (less than high school, 12 years or completed high school, some college, and college graduate or higher). Participants’ health status items include self-reported general health (excellent, very good, good, fair, and poor), ever had cancer (have you ever been diagnosed as having cancer?: yes/no), and depression and anxiety index calculated and categorized using Patient Health Questionnaire-4 [17]. A perceived provider using electronic health (eHealth) records system was assessed as “Do any of your doctors/HCP maintain your medical records in a computerized system? (yes/no).” Trust of a doctor is evaluated by the item “In general, how much would you trust information from a doctor?” (a lot, some, a little, and not at all). The frequency of visiting a nonemergency provider was assessed during the past 12 months (1 time, 2 times, 3 times, 4 times, 5-9 times, 10 or more times).

Statistical Analyses

The study’s analytic goal is to examine the associations of our independent variables of interest (including PPR and its interaction term with time) with WIB across 5 time points. To achieve this goal, we utilized 2 logistic regression models structured as follows:

Model 1 is shown in equation 1 of Figure 1, where year is a continuous variable for time points and PPR represents PPR for respondent *i*. The quadratic term year $_i^2$ represents the curvilinear effect of year on logit (WIB). Model 2 is shown in equation 2 of Figure 1, where $\boldsymbol{X}$ is a vector of k control variables include gender, education, census division, urbanity, age group, occupation, whether born in the United States, general health, provider maintains electronic medical record, depression, trust in doctors, ever had cancer, and frequency to visit providers. The error term ε_i is assumed to be independent of all independent variables and had a mean of zero.

Figure 1. Equations for logistic regression models, slope, and change of point for time trends.

(a)

$$ln\frac{Pr(WIB_i)}{1-Pr(WIB_i)} = \beta_0 + \beta_1 year_i + \beta_2 year_i^2 + \beta_3 PPR_i + \beta_4 year * PPR_i + \varepsilon_i \quad (1)$$

(b)

$$ln\frac{Pr(WIB_i)}{1-Pr(WIB_i)} = \beta_0 + \beta_1 year_i + \beta_2 year_i^2 + \beta_3 PPR_i + \beta_4 year * PPR_i + \beta_k \boldsymbol{X}_{i,k} + \varepsilon_i \quad (2)$$

(c)

$$\frac{dLogit(WIB_i)}{dyear_i} = \beta_1 + 2\beta_2 year_i + \beta_4 PPR_i \quad (3)$$

(d)

$$-\frac{\beta_1 + \beta_4 PPR_i}{2\beta_2} \quad (4)$$

The slope of time trend for logit of WIB is defined as the change of log-odds or logit per year and operationalized as shown in equation 3 of Figure 1, where PPR is treated as a constant. A change of point for the time trend is defined as the time when the slope is 0 and calculated as shown in equation 4 of Figure 1.

Odds ratio (OR) is defined as the exponential of the coefficient $\beta(e^{\beta})$ Multicollinearity was inspected by calculating variance inflation factors (VIFs) for all variables. Findings resulted in VIFs lower than 5 among all independent and covariate variables, suggesting that multicollinearity is not at a concerning level [18]. The area under the receiver operating characteristics curve (C-statistics) was used to assess model performance and suggest sufficient model fit.

Chi-square tests were conducted to examine bivariate associations between outcome and demographic measures. All analyses were conducted using SAS version 9.4 (SAS Institute Inc). All descriptive analyses and logistic regression models were weighted to reflect a nationally representative estimate using the Statistical Analysis System procedures *Proc Surveyfreq* and *Proc Surveylogistic*, respectively. A 2-tailed $P<.05$ was considered statistically significant in this study.

Results

Table 1 shows sample characteristics among study participants, which are stratified by their WIB status. Weighted proportions were calculated through the Jackknife estimation method to generalize findings to the population. Between 2011 and 2014, the proportioned differences between respondents with and without WIB increased from 2.29% to 4.61% and then dropped to −5.38% in 2018, indicating a potential nonlinear time trend of WIB. There are no significant differences between WIB and none-WIB groups when examining the gender and education sample distributions. Compared with respondents who did not withhold information from their providers, respondents who had WIB were more likely to be aged 25 to 44 years (difference: 13.82%) and employed (difference: 10.79%).

The weighted prevalence of WIB in the United States was assessed from 2011 to 2018. As Figure 2 shows, the WIB prevalence has an increasing trend before 2014, which has the highest rate of 13.6% and then decreased after 2014 to 8.6% (Multimedia Appendix 1). When examining the proportions of different levels of PPR across 5 time points, results show that PPR improves after 2011 as shown in Figure 3 by the increasing proportions of respondents with “good” PPR (tertile 3) and decreasing proportions of respondents with “poor” PPR (tertile 1). To explore the moderation effects of PPR on time trends for WIB (Figure 4), we plotted the weighted prevalence of WIB across time points by 3 categorized PPR levels. Results suggest that PPR affects the curvilinear time trend of WIB in 2 ways: on the one hand, the worse PPR is associated with an escalated prevalence of WIB in the year 2011; on the other hand, worse PPR also inverts the sign of the starting slope of curves, which is consistent with prior studies.

Table 1. Sample characteristics of the nationally representative sample across 5 iterations.

Variable	Withhold information, n (%): yes	Withhold information, n (%): no	Difference (%)	*P* value[a]
Year				.003
2011	339 (21.5)	2357 (19.2)	2.29	
2012	293 (21.7)	2018 (19.1)	2.62	
2014	332 (23.2)	2039 (18.6)	4.61	
2017	187 (17.3)	2002 (21.4)	−4.14	
2018	199 (16.3)	2188 (21.7)	−5.38	
Gender				.14
Male	500 (48.9)	4121 (45.8)	3.16	
Female	850 (51.1)	6483 (54.2)	−3.16	
Education				.46
Less than high school	79 (7.9)	691 (8.4)	−0.48	
12 years or completed high school	182 (17.8)	1936 (19.5)	−1.67	
Some college	451 (38.8)	3155 (35.5)	3.27	
College graduate or higher	638 (35.5)	4822 (36.7)	−1.12	
Age group				<.001
18-24	36 (4.8)	294 (10.8)	−5.97	
25-44	462 (46.1)	2521 (32.3)	13.82	
45-64	652 (41.0)	4506 (38.4)	2.6	
65+	200 (8.1)	3283 (18.5)	−10.45	
Occupation				<.001
Employed	838 (68.3)	5476 (57.5)	10.79	
Unemployed	91 (7.0)	444 (5.5)	1.53	
Retired	190 (7.8)	3163 (18.5)	−10.74	
Disabled	121 (7.1)	670 (5.4)	1.76	
Other	110 (9.7)	851 (13.1)	−3.34	

[a]*P* values are calculated from chi-square tests to show association between characteristics variables and WIB.

Figure 2. Unadjusted prevalence trend of withholding information behavior between 2011 and 2018 based on the Health Information National Trends Survey.

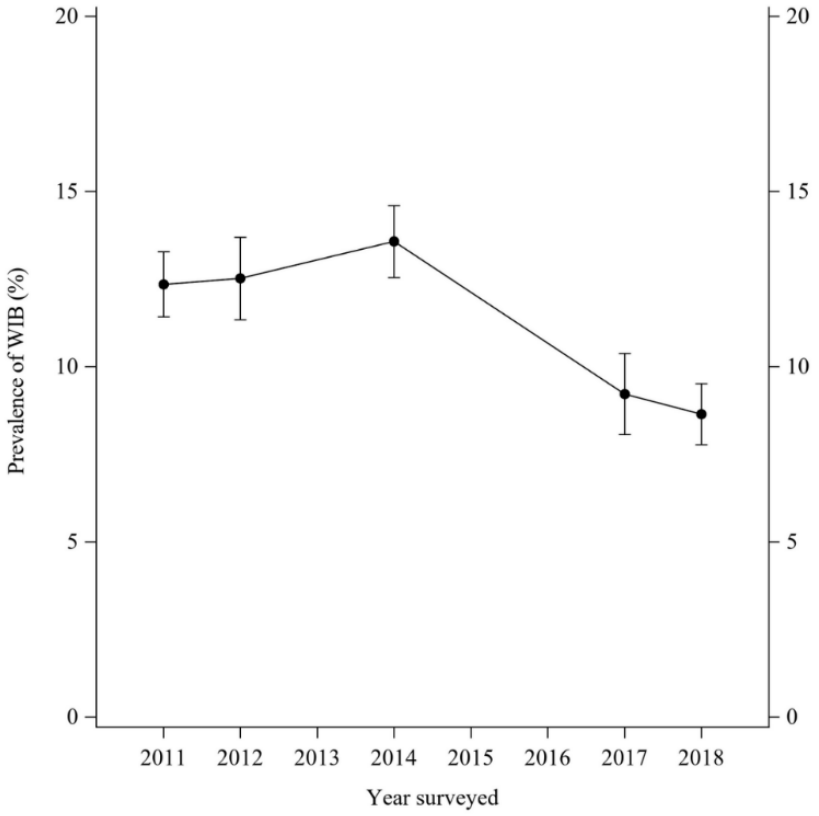


Figure 3. Unadjusted proportions of patient-physician relationship tertiles based on the Health Information National Trends Survey between 2011 and 2018. PPR: patient-physician relationship.

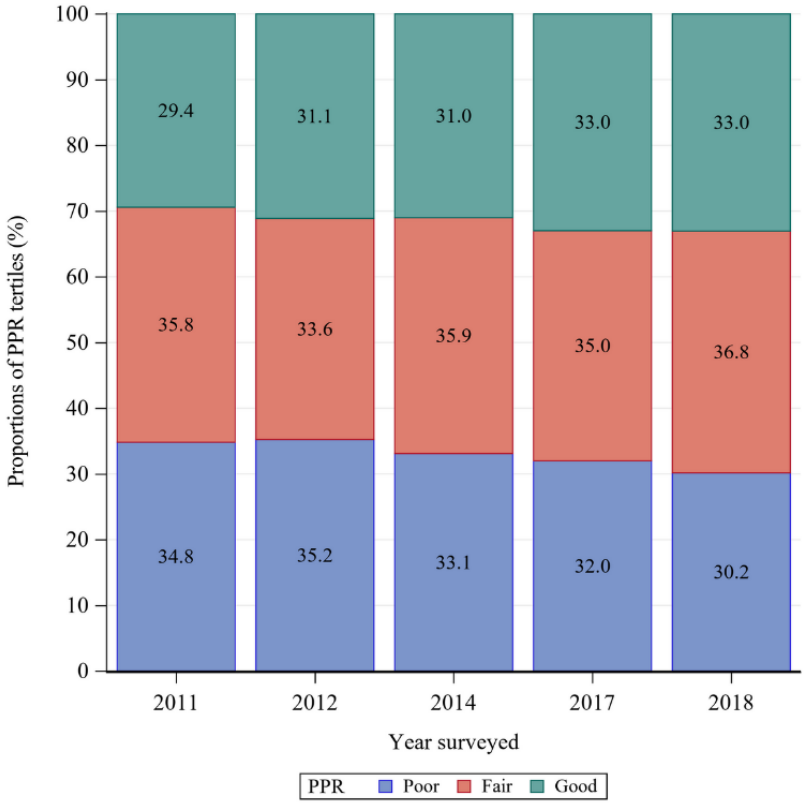
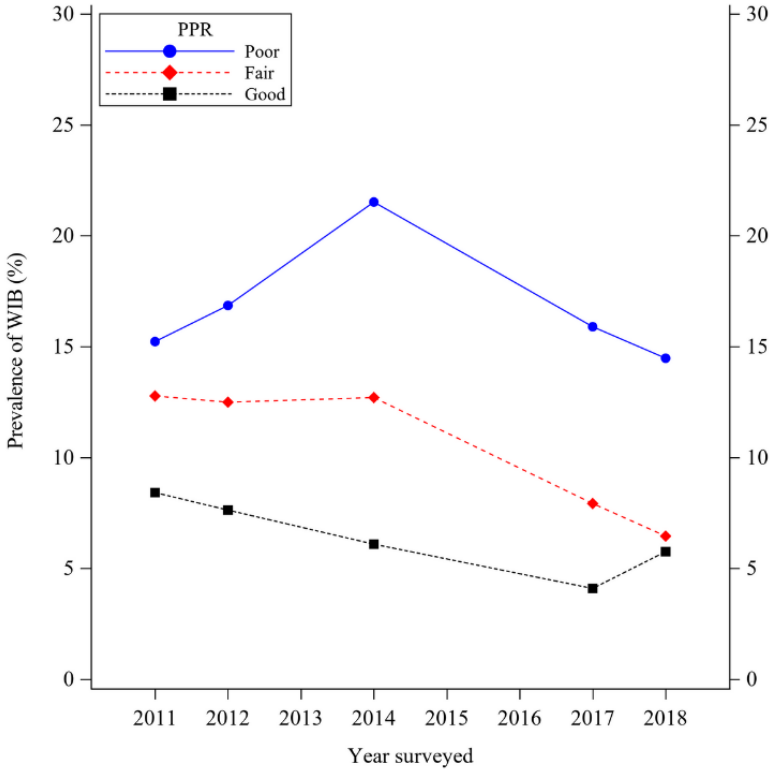


Figure 4. Unadjusted prevalence trends of withholding information behavior grouped by levels of patient-physician relationship and based on responses from the Health Information National Trends Survey between 2011 and 2018. WIB: withholding information behavior.



For the multivariable logistic regression model, results displayed in Table 2 show the model coefficients and statistical significance for linear time trend, quadratic time trend, PPR, and interaction between linear time trend and PPR (model 1 as baseline model). A second logistic regression model (model 2) including covariates such as gender, age, and group education was used to examine whether coefficients were attenuated with the inclusion of confounding variables. The interaction term between quadratic time trend and PPR was not statistically significant and, therefore, removed from the model for simplicity of interpretation. Compared with the baseline model, model with covariates did not change the sign or strength of coefficient estimates, except PPR simple effect was about half of that in the baseline model (beta=−.286 in model 2 vs beta=−.462 in model 1). Accordingly, the odds for every 1 unit of PPR improvement as operationalized in our study is approximately a 24.9% decrease in withholding patterns (OR 0.751, 95% CI 0.598-0.943) in model 2 and 37% decrease (OR 0.63, 95% CI 0.518-0.766) in model 1 in the year 2011. For model 1, when holding PPR constant at 3 (the best PPR), the initial linear time trend slope was 0.052 (OR 1.053) and not significant (*P*=.48, not shown in the table). Due to the significant quadratic term (*P*=.04), the slope decreased by 0.044 per year, resulting in a decreasing trend after year 2012. However, when PPR was worse, for example, PPR=0, the linear time trend slope was significantly higher (beta=.233; SE=0.101; OR 1.263, 95% CI 1.030-1.547) and the change of point occurred later (about 5.3 years after 2011 when PPR=0). Figure 5 demonstrated visualization of the time trend for WIB probability, which was predicted by the logistic regression model 1, against the times surveyed at different PPR levels (0, 1, 2, and 3). PPR significantly modified time trend of WIB through increasing the initial slope in 2011 and, therefore, postponed the time for change of points.

Table 2. Coefficients of multiple logistic regression models with and without covariates.

Model and term	Coefficient estimates (beta)	SE	*P* value	Odds ratio and 95% CI
Model 1: without covariates				
Intercept	−.950[a]	0.226	<.001	0.387 (0.246-0.609)[b]
Year	.233	0.101	.03	1.263 (1.030-1.547)
Year^2	−.022	0.011	.04	0.978 (0.957-0.999)
PPR[c]	−.462	0.097	<.001	0.630 (0.518-0.766)
Year×PPR	−.060	0.025	.02	0.941 (0.896-0.989)
Model 2: with covariates[d]				
Intercept	−1.763	0.58	.004	0.172 (0.053-0.551)
Year	.253	0.104	.02	1.288 (1.044-1.588)
Year^2	−.024	0.011	.04	0.977 (0.955-0.999)
PPR	−.286	0.113	.01	0.751 (0.598-0.943)
Year×PPR	−.068	0.026	.01	0.934 (0.887-0.984)

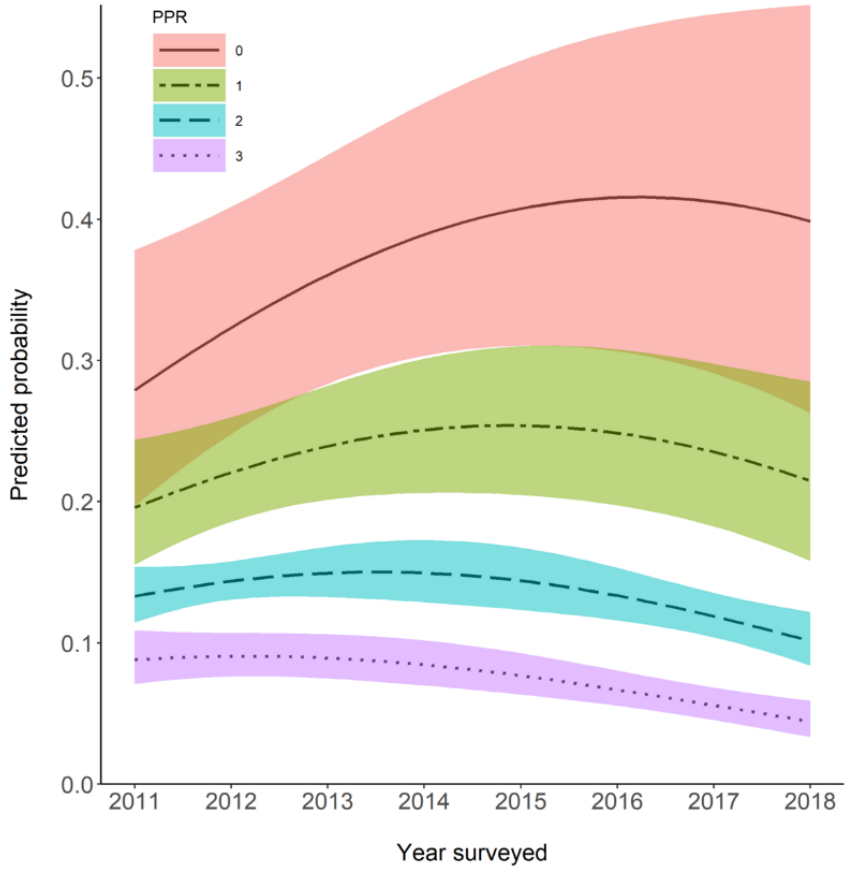
[a]The estimate for the intercept is the baseline log-odds when year is in 2011 and PPR is 0.

[b]The odds ratio and 95% CI for intercept are the baseline odds and 95% CI when the year is in 2011 and patient-physician relationship is 0.

[c]PPR: patient-physician relationship.

[d]Covariates include gender, education, race, urbanity, age group, occupation, census division, born in the United States, general health status, provider maintains electronic medical record, depression, trust doctors, ever had cancer, and frequency to visit providers.

Figure 5. Predicted probability of withholding information behaviors at various patient-physician relationship values. PPR: patient-physician relationship.



Discussion

Principal Findings

Our results suggest complex relationships between PPR and time trends of patients' WIB in the United States. To the best of our knowledge, this study is the first to report the trend of WIB with close to a decade worth of responses (2011-2018). A previous report indicated that the proportion of individuals who withheld information from their providers were slightly decreased from 2012 to 2014, but the change was not statistically significant [10]. Our results revealed that the self-reported prevalence of withholding information first increased during 2011 to 2014 and then dropped to the lowest level in 2018. Although this study is not powered or designed to identify causal factors for the temporary increasing WIB between 2012 and 2014, one has considered whether the timing between this increase with the rise of health care data breaches between 2012 and 2015 [9] is beyond coincidence. Our data showed that PPR has improved steadily since 2011. As indicated by our logistic models, PPR is negatively associated with patients' WIB, which supports our hypothesis that improved PPR lowers odds of patients exhibiting WIB patterns. Therefore, improved PPR may have led to the decreasing trend of WIB after 2014. However, we cannot rule out other unmeasured factors that contribute to the decreasing trend of WIB. Moreover, findings from this study highlight that the curvilinear time trend in WIB was moderated significantly by PPR, where our third focus was shown by the significant interaction term between year and PPR. In the population who reported the best PPR (PPR=3), the WIB prevalence starts to drop continuously as early as in 2012. When PPR was worse, the time for this change of point to occur was postponed to later years. Thus, improving PPR appears to have positive effects on reducing patients' WIB both within and across time points surveyed.

It is not surprising that PPR is associated with patients' WIBs. In Abdelhamid et al's study [7], the PPR was found to be positively associated with the intention to share PHI electronically, which is consistent with our findings that better relationships lead to lower rates of WIB. In a qualitative study from a small sample of Malaysian, doctor's interpersonal and communication skills were reported by all participants to affect their decisions to disclose medicine use information [14]. Another qualitative study among a Latino group found that low-quality relationships diminished participants' willingness to disclose their health information [13]. In addition, existing findings suggest that (1) patients' satisfaction with involvement in health care decisions and (2) perception of their doctors' interest in their general health status also positively influence their decisions to share their eHealth findings [19]. Our study is consistent with these findings on the effects of PPR on patients' disclosure behaviors and extends them to a nationally representative sample across multiple years. Due to the lack of literature in time trends analysis of WIB, our results for the first time reported the moderation effects of PPR on time trends, providing valuable information and insights for policy makers.

Although other factors such as demographics, socioeconomic status, and trust in doctors were not the focus of our study, we observed interesting findings that might be of interest to researchers in this area of study (Multimedia Appendix 2). For example, depression status is significantly associated with WIB in our study, which is consistent with previous observational research [5,20]. Other studies also reported that more depressive symptoms lead to decreased odds of patients disclosing medical information with their doctors [21,22]. Employment status as a dichotomous measure (ie, unemployed vs employed) was not found to be associated with WIB in previous research [5,20]. However, when including more subtypes of employment status, findings suggest that retired individuals were less likely to have WIB compared with individuals reporting as currently employed. This relationship among retired Americans is consistent with our results that older adults have lower odds of WIB than that of middle-aged population. A previous study showed that the top reasons for patients' failures to disclose information are related to trust in clinicians and stigmatization of health behavior [4]. The literature recognizes trust as a factor for improving patients willingness to participate in research and share information [23-27]. The association between WIB and factors such as trust in doctors, PPR, and depression status revealed that these factors may be critical contributors to WIB. Therefore, efforts are warranted for improving relationships between patients and providers and reducing self-stigmatization of patients who have mental diseases to reduce the likelihood of patients' WIB. We did not find any statistically significant interaction effects between PPR and these factors on WIB or the interaction between year and these factors. If consistent in future studies, these findings suggest that providers working with patients at risk for mental health conditions may need to be extra mindful of communicative interactions to reduce odds of WIB patterns.

Limitations and Strength

Similar to other observational studies, this study also has limitations because of the nature of sampling design. First, our analyses were cross-sectional in nature, though we included multiple years of surveys in the trend analysis. The association between predictors and WIBs was not supported by causal inference; therefore, findings should be interpreted carefully and considered as evidence supporting the allocation of resources to examine WIB using a randomized controlled trial research design. Second, HINTS were based on self-reported responses, which are subject to social desirability bias and measurement errors [28]. However, this limitation may be remedied by the stratified random sampling and weighting techniques common in nationally representative complex sampling surveys. Third, our study may be limited by not including various covariates unavailable in the HINTS study. Conversely, a strength of this study is that findings were produced from multiple years of HINTS data to track WIB at a national level. This advantage allowed us to assess moderation effects of PPR on WIB trends, which cannot be detected with limited years of HINTS data [5]. Another strength of our study is that we examined the curvilinear time effects, which allow for a more accurate presentation of WIB over multiple years. Previous research appears overly reliant on statistical methods that assume a linear relationship between continuous predictors and logit of outcome in a logistic regression, which obscures

the curvilinear patterns of predictors and renders their models less representative for actual data [29]. In addition, our analyses on "change of points" integrates measures of interest using polynomial terms, which is suggested to unveil useful information for policy makers and practitioners [30].

Conclusions

We found that there is a curvilinear time trend for the prevalence of patients' WIB. In addition, we found that the PPR is significantly associated with whether patients withhold information from their providers. Moreover, our analyses supported that PPRs moderate the time trends of withholding behavior, and the low quality of relationships between patients and providers postpones the change of point for the decreasing trend. As previously mentioned, the findings from this study in and of itself are not sufficient to motivate changes in national policy. Nevertheless, we believe our findings are interesting enough to warrant further investigation, and if reproducible, the study reinforces the support for interventions bolstering the PPR. Therefore, if findings remain similar in future studies, events and other factors that lower chances of patients fully divulging critical health information can be reduced through improved PPRs. To advance research in this area of study, we believe it to be prudent for future research to replicate this study design with WIB and PPR responses among patients collected on a more granular timescale (eg, "day" and "week"), with inclusion of specific medical breaches to examine moderating effects of medical breaches. The withholding of one's health behavior during medical visits has serious implications on population health. As such, ways to reduce this behavior are of great importance to society.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weighted prevalence and 95% confidence intervals of withholding information behavior across 5 survey years.
[DOCX File , 16 KB - jmir_v22i1e16713_app1.docx]

Multimedia Appendix 2
Odds ratio and 95% confidence intervals of covariates in a multiple logistic regression models (model 2).
[DOCX File , 18 KB - jmir_v22i1e16713_app2.docx]

References

1. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
2. Peterson MC, Holbrook JH, von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med 1992 Feb;156(2):163-165 [FREE Full text] [Medline: 1536065]
3. Ohm F, Vogel D, Sehner S, Wijnen-Meijer M, Harendza S. Details acquired from medical history and patients' experience of empathy--two sides of the same coin. BMC Med Educ 2013 May 9;13:67 [FREE Full text] [doi: 10.1186/1472-6920-13-67] [Medline: 23659369]
4. Levy AG, Scherer AM, Zikmund-Fisher BJ, Larkin K, Barnes GD, Fagerlin A. Prevalence of and factors associated with patient nondisclosure of medically relevant information to clinicians. JAMA Netw Open 2018 Nov 2;1(7):e185293 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.5293] [Medline: 30646397]
5. Walker DM, Johnson T, Ford EW, Huerta TR. Trust Me, I'm a Doctor: examining changes in how privacy concerns affect patient withholding behavior. J Med Internet Res 2017 Jan 4;19(1):e2 [FREE Full text] [doi: 10.2196/jmir.6296] [Medline: 28052843]
6. Patel V, Beckjord E, Moser RP, Hughes P, Hesse BW. The role of health care experience and consumer information efficacy in shaping privacy and security perceptions of medical records: national consumer survey results. JMIR Med Inform 2015 Apr 2;3(2):e14 [FREE Full text] [doi: 10.2196/medinform.3238] [Medline: 25843686]
7. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
8. Agaku IT, Adisa AO, Ayo-Yusuf OA, Connolly GN. Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. J Am Med Inform Assoc 2014;21(2):374-378 [FREE Full text] [doi: 10.1136/amiajnl-2013-002079] [Medline: 23975624]
9. HIPAA Journal. Healthcare Data Breach Statistics URL: https://www.hipaajournal.com/healthcare-data-breach-statistics/ [accessed 2019-05-20]
10. Patel V, Hughes P, Barker W, Moon L. Health IT Dashboard. 2016 Feb. Trends in Individuals Perceptions regarding Privacy and Security of Medical Records and Exchange of Health Information: 2012-2014 URL: https://dashboard.healthit.gov/evaluations/data-briefs/trends-individual-perceptions-privacy-security-ehrs-hie.php [accessed 2019-05-19]

11. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. J Gen Intern Med 1999 Jan;14(Suppl 1):S34-S40 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00263.x] [Medline: 9933493]
12. Larson EB, Yao X. Clinical empathy as emotional labor in the patient-physician relationship. J Am Med Assoc 2005 Mar 2;293(9):1100-1106. [doi: 10.1001/jama.293.9.1100] [Medline: 15741532]
13. Julliard K, Vivar J, Delgado C, Cruz E, Kabak J, Sabers H. What Latina patients don't tell their doctors: a qualitative study. Ann Fam Med 2008;6(6):543-549 [FREE Full text] [doi: 10.1370/afm.912] [Medline: 19001307]
14. Kelak J, Cheah W, Safii R. Patient's decision to disclose the use of traditional and complementary medicine to medical doctor: a descriptive phenomenology study. Evid Based Complement Alternat Med 2018;2018:4735234 [FREE Full text] [doi: 10.1155/2018/4735234] [Medline: 29636778]
15. 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]
16. Hoffman L. Longitudinal Analysis (Multivariate Applications Series). Abingdon-on-Thames: Routledge; 2015.
17. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
18. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
19. de Rosis S, Barsanti S. Patient satisfaction, e-health and the evolution of the patient-general practitioner relationship: evidence from an Italian survey. Health Policy 2016 Nov;120(11):1279-1292 [FREE Full text] [doi: 10.1016/j.healthpol.2016.09.012] [Medline: 27836231]
20. Campos-Castillo C, Anthony D. The double-edged sword of electronic health records: implications for patient disclosure. J Am Med Inform Assoc 2015 Apr;22(e1):e130-e140. [doi: 10.1136/amiajnl-2014-002804] [Medline: 25059953]
21. Beverly EA, Ganda OP, Ritholz MD, Lee Y, Brooks KM, Lewis-Schroeder NF, et al. Look who's (not) talking: diabetic patients' willingness to discuss self-care with physicians. Diabetes Care 2012 Jul;35(7):1466-1472 [FREE Full text] [doi: 10.2337/dc11-2422] [Medline: 22619085]
22. Bell RA, Franks P, Duberstein PR, Epstein RM, Feldman MD, Garcia EF, et al. Suffering in silence: reasons for not disclosing depression in primary care. Ann Fam Med 2011;9(5):439-446 [FREE Full text] [doi: 10.1370/afm.1277] [Medline: 21911763]
23. Tait AR, Voepel-Lewis T, Malviya S. Participation of children in clinical research: factors that influence a parent's decision to consent. Anesthesiology 2003 Oct;99(4):819-825. [doi: 10.1097/00000542-200310000-00012] [Medline: 14508312]
24. Lieberman L, Meana M, Stewart D. Cardiac rehabilitation: gender differences in factors influencing participation. J Womens Health 1998 Aug;7(6):717-723. [doi: 10.1089/jwh.1998.7.717] [Medline: 9718540]
25. Llewellyn-Thomas H, McGreal M, Thiel E, Fine S, Erlichman C. Patients' willingness to enter clinical trials: measuring the association with perceived benefit and preference for decision participation. Soc Sci Med 1991;32(1):35-42. [doi: 10.1016/0277-9536(91)90124-u] [Medline: 2008619]
26. Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. Lancet Oncol 2006 Feb;7(2):141-148. [doi: 10.1016/S1470-2045(06)70576-9] [Medline: 16455478]
27. Keusch F, Rao R, Chang L, Lepkowski J, Reddy P, Choi SW. Participation in clinical research: perspectives of adult patients and parents of pediatric patients undergoing hematopoietic stem cell transplantation. Biol Blood Marrow Transplant 2014 Oct;20(10):1604-1611 [FREE Full text] [doi: 10.1016/j.bbmt.2014.06.020] [Medline: 24972252]
28. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. J Appl Psychol 1999;84(5):754-775. [doi: 10.1037/0021-9010.84.5.754]
29. Babowitch JD, Mitzel LD, Vanable PA, Sweeney SM. Depressive symptoms and condomless sex among men who have sex with men living with HIV: a curvilinear association. Arch Sex Behav 2018 Oct;47(7):2035-2040. [doi: 10.1007/s10508-017-1105-3] [Medline: 29110113]
30. Osborne JW. Best Practices in Logistic Regression. Thousand Oaks: SAGE Publications; 2014.

Abbreviations

eHealth: electronic health
HINTS: Health Information National Trends Survey
OR: odds ratio
PHI: protected health information
PPR: patient-physician relationship
WIB: withholding information behavior
VIF: variance inflation factor

Edited by G Eysenbach; submitted 16.10.19; peer-reviewed by D Walker, KM Kuo, L Rusu; comments to author 04.11.19; revised version received 03.12.19; accepted 16.12.19; published 29.01.20.

Please cite as:
Yang X, Parton J, Lewis D, Yang N, Hudnall M
Effect of Patient-Physician Relationship on Withholding Information Behavior: Analysis of Health Information National Trends Survey (2011-2018) Data
J Med Internet Res 2020;22(1):e16713
URL: http://www.jmir.org/2020/1/e16713/
doi:10.2196/16713
PMID:32012083

Review

Measures of Effectiveness, Efficiency, and Quality of Telemedicine in the Management of Alcohol Abuse, Addiction, and Rehabilitation: Systematic Review

Clemens Scott Kruse[1], MHA, MSIT, MBA, PhD; Kimberly Lee[1], MHA, PhD; Jeress B Watson[1], BA, MHA; Lorraine G Lobo[1], BS, MHA; Ashton G Stoppelmoor[1], BA, MHA; Sabrina E Oyibo[1], BA, MHA

School of Health Administration, Texas State University, San Marcos, TX, United States

Corresponding Author:
Clemens Scott Kruse, MHA, MSIT, MBA, PhD
School of Health Administration
Texas State University
601 University Drive
Encino Hall, Room 250
San Marcos, TX, 78666
United States
Phone: 1 2103554743
Fax: 1 5122458712
Email: scottkruse@txstate.edu

Abstract

Background: More than 18 million Americans are currently suffering from alcohol use disorder (AUD): a compulsive behavior of alcohol use as a result of a chronic, relapsing brain disease. With alcohol-related injuries being one of the leading causes of preventable deaths, there is a dire need to find ways to assist those suffering from alcohol dependence. There still exists a gap in knowledge as to the potential of telemedicine in improving health outcomes for those patients suffering from AUD.

Objective: The purpose of this systematic review was to evaluate the measures of effectiveness, efficiency, and quality that result from the utilization of telemedicine in the management of alcohol abuse, addiction, and rehabilitation.

Methods: This review was conducted utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. The articles used in this analysis were gathered using keywords inclusive of both *telemedicine* and *alcohol abuse*, which were then searched in the Cumulative Index to Nursing and Allied Health Literature, Cochrane, and MEDLINE (PubMed) databases. A total of 22 articles were chosen for analysis.

Results: The results indicated that telemedicine reduced alcohol consumption. Other common outcomes included reduced depression (4/35, 11%), increased patient satisfaction (3/35, 9%), increase in accessibility (3/35, 9%), increased quality of life (2/35, 6%), and decreased cost (1/35, 3%). Interventions included mobile health (11/22, 50%), electronic health (6/22, 27%), telephone (3/33, 14%), and 2-way video (2/22, 9%). Studies were conducted in 3 regions: the United States (13/22, 59%), the European Union (8/22, 36%), and Australia (1/22, 5%).

Conclusions: Telemedicine was found to be an effective tool in reducing alcohol consumption and increasing patients' accessibility to health care services or health providers. The group of articles for analysis suggested that telemedicine may be effective in reducing health care costs and improving the patient's quality of life. Although telemedicine shows promise as an effective way to manage alcohol-related disorders, it should be further investigated before implementation.

(J Med Internet Res 2020;22(1):e13252) doi:10.2196/13252

KEYWORDS

telemedicine; telehealth; mHealth; alcohol abuse; rehabilitation; alcohol use disorder

Introduction

Background

More than 18 million Americans are currently suffering from alcohol use disorder (AUD) [1]. The National Institute on Alcohol Abuse and Alcoholism defines AUD as a *chronic relapsing brain disease in which a person or individual displays compulsive alcohol use, loss of control in regards to alcohol intake, and a negative emotional state when not using* [2]. An individual's consistent engagement in the use of alcohol may prove harmful not only to the individual's health but to the health of others as well.

A growing concern in the health care field is the alarming number of preventable deaths that result from alcohol-related injuries. It is estimated that excessive alcohol use is responsible for approximately 88,000 deaths and approximately 2.5 million potential life years lost annually in the United States and in the European Union; approximately 1 in 4 deaths of males aged 15 to 39 years are because of alcohol [3,4]. Owing to the substantial health and economic barriers that result from alcohol dependence, there exists a dire need for an alternative and innovative solution for managing alcohol abuse, addiction, and rehabilitation.

Telemedicine, specifically videoconferencing, is identified as a successful tool in reducing the effects of adults suffering from AUD, but videoconferencing is only one small aspect of telemedicine that can offer assistance to this condition [5]. Telemedicine, as defined by the World Health Organization is:

The delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities [6].

The use of telemedicine in treating alcohol dependence helps increase the patient's access to providers as well as increasing social support outside of the health care setting [5]. Both factors have demonstrated effectiveness in managing addiction and/or preventing relapse. Telemedicine has the ability to reduce obstacles such as geographical locations and time while still continuing to deliver the same (or better) quality health care [5]. Despite these findings, there still is a gap in knowledge as to the potential of telemedicine-based treatments in improving health outcomes in AUD patients.

A review was conducted on this topic in 2012 [7]. It provided an extensive review of literature (n=50); studies were conducted in 7 countries; interventions reported were telephone and voice response, videoconferencing, text messaging, Web, email, and chat; and outcomes reported were participation, substance use, satisfaction, and resource utilization. However, technology has evolved a great deal in telecommunications since 2012, and some researchers suggest reviews should be repeated after 2 years [8]. Another review was conducted more recently in Australia [9]. It only analyzed 19 articles and focused purely on mobile apps as an intervention in 1 country [9].

Objective

The purpose of this systematic review was to analyze and evaluate current literature in regard to the measures of effectiveness, efficiency, and quality that result from the use of telemedicine in managing alcohol abuse, addiction, and rehabilitation in patients. Effectiveness was measured through outcomes. Efficiency was measured through cost. Quality was measured in terms of safety, timeliness, access, patient satisfaction, or quality of life.

Methods

Protocol and Registration

The research process was structured following the Kruse Protocol for Writing Systematic Reviews [10] and the Assessment for Multiple Systematic Reviews (AMSTAR) [11]. Findings were reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [12]. The PRISMA checklist for this review is listed in Multimedia Appendix 1. The authors started the process with the identification of a specific set of key terms intended to produce a list of inclusive articles from a variety of domains that were directly related to the topic of discussion. Both *telemedicine* and *alcohol abuse* were searched on MEDLINE (PubMed) using the US National Library of Medicine's Medical Subject Headings (MeSH). A total of 5 subheadings were identified under the search of *telemedicine*, and 13 subheadings were identified for *alcohol abuse*. For the purposes of this systematic review, these 18 subheadings were selected as the final set of key terms to be used for searches within the research databases. To ensure we captured at least 10 years of research, searches were conducted on December 31, 2018, and the time period of inclusion was December 1, 2009 to December 31, 2018.

The subheadings were separated utilizing Boolean operators to ensure the search produced articles that are inclusive of both alcohol abuse and telemedicine or terms related to those topics. See Multimedia Appendix 2 for the exact Boolean phrase used in all databases.

Eligibility Criteria, Information Sources, and Search

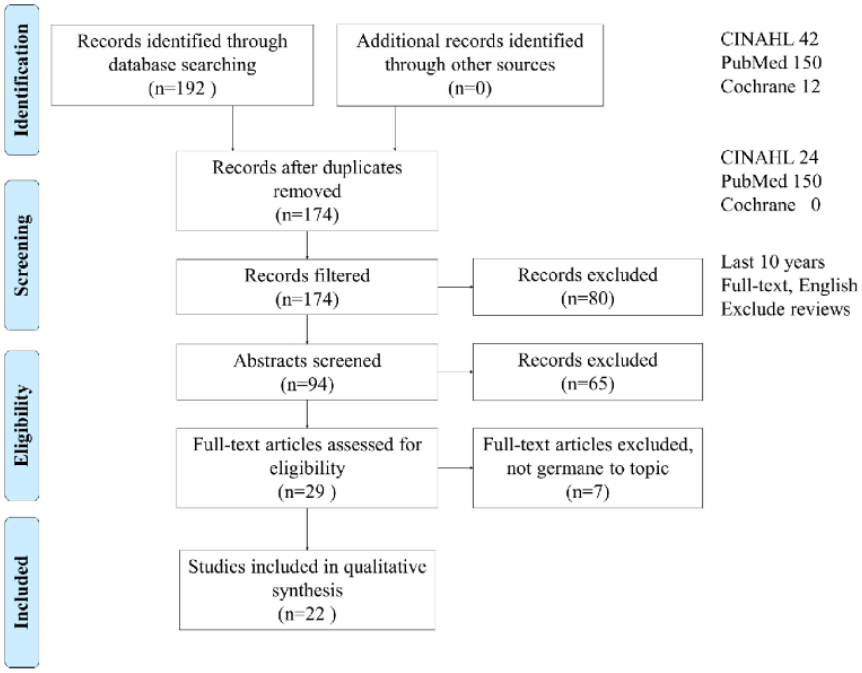
A total of 3 research databases were chosen: Cumulative Index to Nursing and Allied Health Literature (CINAHL), PubMed (MEDLINE), and Cochrane. These databases were chosen because of their wide availability, following the Kruse Protocol for writing a systematic review, and because they are recommended by the National Institutes of Health [13].

Articles were eligible for analysis if they were published in the last 10 years and included information on the use of telemedicine for the treatment of AUD. We preferred the articles also included data on outcomes, but that was not an eligibility criterion.

Using the key terms from MeSH, the initial search in the CINAHL yielded a total of 42 articles, the PubMed database yielded 150 articles, and Cochrane yielded 12. The search process and criteria are visually represented in Figure 1. To limit the search to recent and relevant articles, filters were applied to both CINAHL, PubMed, and Cochrane database

searches. In CINAHL, MEDLINE was excluded to help eliminate duplicates. Filters were applied to both meet our acceptance criteria of the last 10 years and to ensure articles were available (full text, English, no reviews). A total of 94 articles remained for the next step in the process.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. CINAHL: Cumulative Index of Nursing and Allied Health Literature.



Study Selection and Data Collection Process

The next step of the systematic review process consisted of a screening of abstracts by the authors. The 94 abstracts were divided between 5 authors to be carefully screened according to the objective statement. Each abstract was screened by at least two authors, and a final consensus meeting was held to finalize the selection of articles to be included in the systematic review. During this process, a total of 65 articles were excluded for not being germane, leaving a total of 29 articles.

The 29 articles were read in their entirety by at least two reviewers and excluded if articles were not related to the objective statement, were study protocols, or were systematic reviews. Similar to the screening process, articles were divided among the 6 reviewers for analysis commensurate with our objective statement. Workload was divided among 6 authors following the Kruse Protocol [10]: JW, LL, AS, and SO each analyzed 10 articles, KL analyzed 5 articles, and CSK analyzed all 29 articles. This process left a total of 22 articles for inclusion in the group for analysis. Figure 1 illustrates the screening and selection process.

Level of Agreement

A kappa statistic was calculated to assess the level of agreement between the authors. The calculated kappa score was 0.85 indicating strong agreement on which articles should be included in the group for analysis [14-16]. The closer this score is to 1, the stronger the agreement.

Data Items, Summary Measures, Synthesis of Results, Bias, and Additional Analysis

During the analysis phase, data items were extracted such as participants, intervention, comparison method, outcomes, as well as measures of effectiveness, efficiency, and quality. Each article was analyzed by at least two reviewers, and independent observations were made. A consensus meeting was held to discuss observations and distill observations to themes. On the basis of the combined set of observations and identified themes, a second read of articles was conducted and another consensus meeting held to ensure an exhaustive data collection. Themes were analyzed for trends. Affinity matrices were created to describe frequencies of data. The frequency of occurrence was recorded for all articles. Multiple articles mentioned multiple themes. Observations of potential bias were recorded.

Results

Study Selection, Study Characteristics, Results of Studies, and Synthesis of Results

Multimedia Appendix 3 lists a detailed summary of the analysis of the 22 articles in the review.

All articles in the analysis discuss telemedicine for the treatment of AUD [17-38].

Additional Analysis

After performing a punctilious data extraction process, the authors were able to identify several potential outcomes that were used to measure the effectiveness of telemedicine in the management of alcohol use, addiction, and rehabilitation: reduction in alcohol consumption, increase in patient

satisfaction, reduction in symptoms of depression, increased accessibility, and decreased cost. The authors selected a total of 7 outcomes as evaluation criteria for article analysis in a total of 35 occurrences (number of times the theme was addressed). Articles can include multiple themes. An affinity matrix of these themes is listed in Table 1.

Out of the outcomes identified, a total of 16 out of 35 (46%) occurrences mentioned that there was a statistically significant reduction in the participants' alcohol consumption [17,18,20-30,32,35,38]. A total of 6 of 35 (17%) showed an increase in cognition [19,20,31,33,36]. A total of 4 of 35 (11%) showed a decrease in depressive symptoms [17,21,26,27]; and 2 themes were each mentioned 3 of 35 (9%) occurrences: increased patient satisfaction [18,22,25] and increased accessibility [17,20,24]. The latter included easily accessible tools for the participants to use that were not associated with the negative stigma of being treated for alcohol abuse. In all, 2 of 35 (6%) occurrences were increased quality of life [20,26], and 1 of 35 (3%) was cost-effectiveness of the intervention [37].

A total of 4 technological interventions were identified (summarized in Table 2). The intervention most often identified in the literature was a mobile app or SMS, which are text messages. This intervention was identified in 11 of the 22 articles (50%) [18,19,23-25,29,31,36]. The next most identified intervention was Web-based (electronic health, eHealth), which occurred in 6 of 22 articles (27%) [17,20,26,37]. One intervention occurred in 3 of the 22 articles (14%): phone-based (voice) [22,35,38]. Finally, 2-way video occurred in 2 of 22 articles (9%) [21,30].

Researchers from 6 different countries generated articles (summarized in Table 3). Researchers in the United States generated more than half the articles, producing 13 of 22 articles (59%) [19-23,28-32,35,38]. Researchers from European countries produced 8 articles including Sweden [24,26], Portugal [33,34], the United Kingdom [18], Spain [25], Germany [17], and the Netherlands [37]. Research from Australia occurred once in the literature [27].

Telemedicine showed positive medical outcomes (effectiveness) in 77% (17/22) of the articles analyzed and resulted in no statistical significance in improvement in the remaining 23% (5/22). There were no observations where telemedicine resulted in a decrease in medical outcome or effectiveness.

Table 1. Themes observed in the literature (N=35).

Themes	Studies (references)	Frequency, n (%)
Reduced alcohol consumption	[17,18,20-30,32,35,38]	16 (46)
Increased cognition	[19,20,31,33,36]	6 (17)
Reduced depression	[17,21,26,27]	4 (11)
Increased patient satisfaction	[18,22,25]	3 (9)
Increased accessibility	[17,20,24]	3 (9)
Increased quality of life	[20,26]	2 (6)
Decreased cost	[37]	1 (3)

Table 2. Interventions observed in the literature (N=22).

Intervention	Studies (references)	Frequency, n (%)
Mobile health	[18,19,23-25,29,31,36]	11 (50)
Electronic health	[17,20,26,37]	6 (27)
Telephone	[22,35,38]	3 (14)
2-way video	[21,30]	2 (9)

Table 3. Summary of country of origin of research (N=22).

Country	Studies (references)	Frequency, n (%)
The United States	[19-23,28-32,35,38]	13 (59)
European Union	[17,18,24-26,33,37]	8 (36)
Australia	[27]	1 (5)

Discussion

Summary of Evidence

The key intent in this systematic review was to identify measures of effectiveness, efficiency, and quality of telemedicine in managing conditions of AUD or some sort of substance dependence. A total of 7 measures of effectiveness and 4 categories of technological intervention from 8 countries were identified in 22 articles. The intervention of technology was associated with positive medical outcomes in 17 of the 22 articles [17-21,23-27,29-33,35,36]. In the other 5 articles, it

equaled the outcome of the control group [22,28,34,37,38]. Most importantly, technological interventions enabled patients to meet their goals with a tool that could be used independent of the health care facility, without the negative stigma associated with alcohol and drug abuse or addition and depressive or posttraumatic stress disorder symptoms.

Subjects, including patients, clients, students, or employees, reported high levels of satisfaction with the technological interventions for several reasons. The technology solutions enabled a variety of positive aspects of care: self-management of subjects' symptoms or conditions, the technology solutions were self-paced, the technology solutions were both synchronous and asynchronous with providers, and technology solutions were based on subjects' preferences. Individuals did not have to travel to the provider or counselor, reducing costs, and addressing concerns over a stigma of being treated for AUD [26]. Subjects accessed solutions via mobile health and eHealth around the clock, 24×7, rather than spending valuable time making and traveling to and from onsite appointments. The telehealth solutions enabled real-time management of temptations and other problems through interactive and often customizable feedback to help patients change their behavior.

Beyond the scope of this study and yet an important consideration, providers incurred costs for the technological interventions reviewed in this study, including initial development, capital securement, supportive programming, administration, personnel, and time and resources for ongoing delivery and aftercare. However, organizations' ability to treat and provide valuable resources to recipients of care outside the brick and mortar far exceeded the cost of the intervention. Organizations were able to expand practice without expanding square footage, and eHealth solutions continued to provide care outside the boundaries of 8 am to 5 pm, a traditional treatment day. An organization with the treatment of alcohol and drug abuse or addiction and depressive symptoms should strongly consider a technological intervention because it would enable the organization to meet this strategic goal without another product line, additional provider staff, or clinic space.

The implementation of telemedicine as an alcohol management treatment option has demonstrated great promise in reducing alcohol consumption and positively affecting other factors associated with alcohol use and improved patient outcomes. Alcohol management is the treatment of alcohol abuse through moderation. It addresses the highly complex and personal issue that alcoholism can create [39]. Alcohol management helps an individual drink in moderation when abstinence programs are ineffective.

The 22 articles analyzed included participants from several countries. The authors identified a total of 7 characteristics for use in measuring telemedicine's effectiveness in AUD patients. Most of the articles mentioned a decrease in alcohol consumption (16/22) and an increase in accessibility (3/22). A social stigma was mentioned that is a cause for concern, and it supports the adoption of telemedicine [26]. The fear is being seen in or around the clinic that treats AUD. Treatment through telemedicine overcomes this stigma by enabling patients to be treated virtually. Other barriers were listed within the articles but were not listed enough times to be able to state that a correlation exists between those barriers and the adoption of telemedicine in treating AUDs.

It is unfortunate that so few articles qualified for analysis based on our selection criteria. To do this study over again, we would expand our selection criteria to enable a larger number of articles to be included in the analysis. The larger number of articles would strengthen the results and associated conclusions and give us a stronger indication of external validity.

Of interesting note, the countries publishing articles on AUD are developed countries. It is unlikely that developing countries do not experience difficulties with AUD. Instead, it is more likely that these countries are not devoting resources to research the topic.

Strategic Leadership and Cultural Implications

Our systematic review illuminated the need for strategic leadership in planning care environments to include innovative, effective, and efficient solutions such as telemedicine. Leaders and experts in the field of health care emphasize the necessity of delivering value-based care according to trends as identified by over 2000 health care leaders in a 2017 survey [40]. Chief executives of hospitals and other care organizations anticipate that consumers will increasingly expect autonomous, user-friendly, and empathetic care, and consumers' expectations will coincide with the necessity of organizations to deliver value, equity, safety, transformation, affordability, innovation, and efficiency [40]. A consistency of subjects experiencing enhanced autonomy and satisfaction occurred in this systematic review, supporting predicted trends in consumers' expectations [17,18,20-22,25-27].

One of the increasingly frequent realities, the need for cybersecurity, will create a culture of increased trust with consumers and will also serve to address one of the cited barriers to traditional care in this study, threat of social exposure and lack of confidentiality around participation in electronic telehealth around a socially sensitive condition, alcohol misuse, abuse, and hazardous lifestyles [26,40]. Increased leadership strategy around creating discipline in behaviors, practices, and process improvement with the intent to protect private data will serve to strengthen the organizational culture and hygiene.

Strategic leaders considering future trends also set a vision of redesigned, decentralized, nontraditional, and remote health care facilities to better meet the needs of consumers, better protect patients and staff, and better prepare for sustainability with greater unpredictability of reimbursement and regulations [40].

Organizational leaders also have the responsibility to incorporate telehealth interventions to not only better reach individuals afflicted by AUDs but also to reduce the societal health burden. To this point, the World Health Organization in its Sustainable Development Goal Agenda of 2030 cites the harmful use of alcohol as a leading risk factor for worldwide health and includes expectations of organizations to assess and implement cost-efficient and multimedia approaches to impact the global issue [41].

Limitations of the Study

The limited number of articles selected in the group for analysis provided was a limitation. With smaller samples, it is possible that the results found within each study cannot be broadly applied to the population. In addition, self-reporting was heavily relied on within these studies and ultimately could have led to self-report bias when patients reported back to the researchers.

Strengths of Our Review

A strength to be considered is the process used to select the final 22 articles for the review. Utilizing the US National Library of Medicine's MeSH enabled an exhaustive search of the literature. Adopting the technique from AMSTAR to use multiple, independent reviewers for both screening abstracts and analyzing articles reduces selection bias [11]. Furthermore, a composite kappa score of 0.85 reflected strong agreement among the authors on their choice of articles based on the objective statement [16]. The kappa score is a measure of agreement between observers [14]. The score establishes confidence in the articles chosen and their relation to the objective statement as there were no articles included in the review that were not agreed upon by 2 or more of the authors. There was also agreement among the authors of the articles chosen that further research is needed to identify those interventions with the strongest level of effectiveness.

Conclusions

Overall, there are many potential applications for telemedicine in improving patient outcomes and strategy in the delivery of care across diverse environments. Telemedicine has the ability to expand the scope of health care by allowing patients the ability to connect to health care providers without the barriers of proximity, without the negative stigma associated with alcohol or drug addiction or mental health, and without the expense of driving to the clinic for an in-person visit.

Telemedicine can allow patients greater access to their provider in times of need. One of the key necessities for successful alcoholism recovery is behavioral therapy and counseling for the patient [39]. The concept used is that the increased communication between the patient and provider will aid the patient in overcoming the barriers associated with alcohol recovery and in turn reduce the number of alcohol-related deaths. With further investigation and research, the use of telemedicine in managing AUDs may be strategically achievable soon.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[DOC File , 64 KB - jmir_v22i1e13252_app1.doc]

Multimedia Appendix 2
Boolean search string.
[DOCX File , 19 KB - jmir_v22i1e13252_app2.docx]

Multimedia Appendix 3
Summary of evidence.
[DOCX File , 29 KB - jmir_v22i1e13252_app3.docx]

References

1. MedlinePlus. Alcohol Use Disorder (AUD) URL: https://medlineplus.gov/alcoholusedisorderaud.html [accessed 2019-09-26]
2. National Institute on Alcohol Abuse and Alcoholism (NIAAA). Alcohol Use Disorder URL: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-use-disorders [accessed 2019-09-26]
3. Centers for Disease Control and Prevention. Fact Sheets - Alcohol Use and Your Health URL: https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm [accessed 2019-09-26]
4. World Health Organization. Alcohol Use URL: http://www.euro.who.int/en/health-topics/disease-prevention/alcohol-use [accessed 2019-09-27]
5. Frueh BC, Henderson S, Myrick H. Telehealth service delivery for persons with alcoholism. J Telemed Telecare 2005;11(7):372-375. [doi: 10.1258/135763305774472060] [Medline: 16238840]
6. World Health Organization. Telemedicine: Opportunities and Developments in Member States. Geneva: World Health Organization; 2010.
7. Young LB. Telemedicine interventions for substance-use disorder: a literature review. J Telemed Telecare 2012 Jan;18(1):47-53. [doi: 10.1258/jtt.2011.110608] [Medline: 22101610]
8. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, Panel for updating guidance for systematic reviews (PUGs). When and how to update systematic reviews: consensus and checklist. Br Med J 2016 Jul 20;354:i3507 [FREE Full text] [doi: 10.1136/bmj.i3507] [Medline: 27443385]

9. Choo CC, Burton AA. Mobile phone apps for behavioral interventions for at-risk drinkers in Australia: literature review. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e18 [FREE Full text] [doi: 10.2196/mhealth.6832] [Medline: 29439946]
10. Kruse CS. Writing a systematic review for publication in a health-related degree program. JMIR Res Protoc 2019 Oct 14;8(10):e15490 [FREE Full text] [doi: 10.2196/15490] [Medline: 31527018]
11. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
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. National Institutes of Health. Systematic Review Service URL: https://www.nihlibrary.nih.gov/services/systematic-review-service [accessed 2019-09-27]
14. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
15. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 1971;76(5):365-377. [doi: 10.1037/h0031643]
16. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
17. Boß L, Lehr D, Schaub MP, Castro RP, Riper H, Berking M, et al. Efficacy of a web-based intervention with and without guidance for employees with risky drinking: results of a three-arm randomized controlled trial. Addiction 2018 Apr;113(4):635-646 [FREE Full text] [doi: 10.1111/add.14085] [Medline: 29105879]
18. Attwood S, Parke H, Larsen J, Morton KL. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. BMC Public Health 2017 May 17;17(1):394 [FREE Full text] [doi: 10.1186/s12889-017-4358-9] [Medline: 28511698]
19. 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 [FREE Full text] [doi: 10.1016/j.jsat.2017.03.011] [Medline: 28476273]
20. 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 Mar;48(2):262-276 [FREE Full text] [doi: 10.1016/j.beth.2016.09.001] [Medline: 28270335]
21. Jaconis M, Santa Ana EJ, Killeen TK, Badour CL, Back SE. Concurrent treatment of PTSD and alcohol use disorder via telehealth in a female Iraq veteran. Am J Addict 2017 Mar;26(2):112-114 [FREE Full text] [doi: 10.1111/ajad.12481] [Medline: 28118514]
22. 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]
23. Muench F, van Stolk-Cooke K, Kuerbis A, Stadler G, Baumel A, Shao S, et al. A randomized controlled pilot trial of different mobile messaging interventions for problem drinking compared to weekly drink tracking. PLoS One 2017;12(2):e0167900 [FREE Full text] [doi: 10.1371/journal.pone.0167900] [Medline: 28146560]
24. Gajecki M, Andersson C, Rosendahl I, Sinadinovic K, Fredriksson M, Berman AH. Skills training via smartphone app for university students with excessive alcohol consumption: A randomized controlled trial. Int J Behav Med 2017 Oct;24(5):778-788 [FREE Full text] [doi: 10.1007/s12529-016-9629-9] [Medline: 28224445]
25. Barrio P, Ortega L, López H, Gual A. Self-management and shared decision-making in alcohol dependence via a mobile app: a pilot study. Int J Behav Med 2017 Oct;24(5):722-727. [doi: 10.1007/s12529-017-9643-6] [Medline: 28236288]
26. Johansson M, Sinadinovic K, Hammarberg A, Sundström C, Hermansson U, Andreasson S, et al. Web-based self-help for problematic alcohol use: a large naturalistic study. Int J Behav Med 2017 Oct;24(5):749-759 [FREE Full text] [doi: 10.1007/s12529-016-9618-z] [Medline: 27900733]
27. Deady M, Mills KL, Teesson M, Kay-Lambkin F. An online intervention for co-occurring depression and problematic alcohol use in young people: Primary outcomes from a randomized controlled trial. J Med Internet Res 2016 Mar 23;18(3):e71 [FREE Full text] [doi: 10.2196/jmir.5178] [Medline: 27009465]
28. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. Am J Psychiatry 2014 Jun;171(6):683-690 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13081055] [Medline: 24700332]
29. Stoner SA, Arenella PB, Hendershot CS. Randomized controlled trial of a mobile phone intervention for improving adherence to naltrexone for alcohol use disorders. PLoS One 2015;10(4):e0124613 [FREE Full text] [doi: 10.1371/journal.pone.0124613] [Medline: 25909320]
30. Staton-Tindall M, Havens JR, Webster JM, Leukefeld C. METelemedicine: a pilot study with rural alcohol users on community supervision. J Rural Health 2014;30(4):422-432 [FREE Full text] [doi: 10.1111/jrh.12076] [Medline: 24800699]

31. Chih M. Exploring the use patterns of a mobile health application for alcohol addiction before the initial lapse after detoxification. AMIA Annu Symp Proc 2014;2014:385-394 [FREE Full text] [Medline: 25954342]
32. Chih M, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat 2014 Jan;46(1):29-35 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]
33. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Caçoete C, et al. Cognitive stimulation through mHealth-based program for patients with alcohol dependence syndrome: a randomized controlled study. J Pain Manag 2016;9(3):235-241 [FREE Full text]
34. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, et al. Executive functioning in alcoholics following an mHealth cognitive stimulation program: randomized controlled trial. J Med Internet Res 2014 Apr 17;16(4):e102 [FREE Full text] [doi: 10.2196/jmir.2923] [Medline: 24742381]
35. Rose GL, Skelly JM, Badger GJ, Naylor MR, Helzer JE. Interactive voice response for relapse prevention following cognitive-behavioral therapy for alcohol use disorders: a pilot study. Psychol Serv 2012 May;9(2):174-184 [FREE Full text] [doi: 10.1037/a0027606] [Medline: 22662731]
36. McTavish FM, Chih M, Shah D, Gustafson DH. How patients recovering from alcoholism use a smartphone intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]
37. Smit F, Lokkerbol J, Riper H, Majo MC, Boon B, Blankers M. Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. J Med Internet Res 2011 Aug 11;13(3):e56 [FREE Full text] [doi: 10.2196/jmir.1694] [Medline: 21840836]
38. McKay JR, van Horn D, Oslin DW, Ivey M, Drapkin ML, Coviello DM, et al. Extended telephone-based continuing care for alcohol dependence: 24-month outcomes and subgroup analyses. Addiction 2011 Oct;106(10):1760-1769 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03483.x] [Medline: 21545667]
39. Alcohol | American Addiction Centers. URL: https://www.alcohol.org/ [accessed 2019-09-26]
40. Society for Healthcare Strategy and Market Development. FUTURESCAN: Moving Ahead on the Long Road Toward Value-Based Healthcare URL: https://www.shsmd.org/education/webcasts/moving-ahead-to-value-based-healthcare [accessed 2019-09-26]
41. World Health Organization. Global Status Report on Alcohol and Health 2018. Geneva: World Health Organization; 2018.

Abbreviations

AMSTAR: Assessment for Multiple Systematic Reviews
AUD: alcohol use disorder
CINAHL: Cumulative Index to Nursing and Allied Health Literature
eHealth: electronic health
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 31.12.18; peer-reviewed by S Shah, S Stoner, D Hacking, Y Yang; comments to author 27.04.19; revised version received 13.06.19; accepted 29.11.19; published 31.01.20.

Please cite as:
Kruse CS, Lee K, Watson JB, Lobo LG, Stoppelmoor AG, Oyibo SE
Measures of Effectiveness, Efficiency, and Quality of Telemedicine in the Management of Alcohol Abuse, Addiction, and Rehabilitation: Systematic Review
J Med Internet Res 2020;22(1):e13252
URL: http://www.jmir.org/2020/1/e13252/
doi:10.2196/13252
PMID:32012048

Original Paper

Classification and Regression Tree and Computer Adaptive Testing in Cardiac Rehabilitation: Instrument Validation Study

Linda Peute[1], BSc, MSc, PhD; Thom Scheeve[2], BSc, MSc; Monique Jaspers[1], BSc, MSc, PhD

[1]Center of Human Factors Engineering of Health Information Technology, Department of Medical Informatics, Amsterdam Institute of Public Health, Amsterdam University Medical Centers, Amsterdam, Netherlands

[2]Signal Processing Systems, Video Coding and Architectures, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Monique Jaspers, BSc, MSc, PhD
Center of Human Factors Engineering of Health Information Technology, Department of Medical Informatics
Amsterdam Institute of Public Health
Amsterdam University Medical Centers
J1b-116
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 640120422
Email: m.w.jaspers@amsterdamumc.nl

Abstract

Background: There is a need for shorter-length assessments that capture patient questionnaire data while attaining high data quality without an undue response burden on patients. Computerized adaptive testing (CAT) and classification and regression tree (CART) methods have the potential to meet these needs and can offer attractive options to shorten questionnaire lengths.

Objective: The objective of this study was to test whether CAT or CART was best suited to reduce the number of questionnaire items in multiple domains (eg, anxiety, depression, quality of life, and social support) used for a needs assessment procedure (NAP) within the field of cardiac rehabilitation (CR) without the loss of data quality.

Methods: NAP data of 2837 CR patients from a multicenter Cardiac Rehabilitation Decision Support System (CARDSS) Web-based program was used. Patients used a Web-based portal, MyCARDSS, to provide their data. CAT and CART were assessed based on their performances in shortening the NAP procedure and in terms of sensitivity and specificity.

Results: With CAT and CART, an overall reduction of 36% and 72% of NAP questionnaire length, respectively, was achieved, with a mean sensitivity and specificity of 0.765 and 0.817 for CAT, 0.777 and 0.877 for classification trees, and 0.743 and 0.40 for regression trees, respectively.

Conclusions: Both CAT and CART can be used to shorten the questionnaires of the NAP used within the field of CR. CART, however, showed the best performance, with a twice as large overall decrease in the number of questionnaire items of the NAP compared to CAT and the highest sensitivity and specificity. To our knowledge, our study is the first to assess the differences in performance between CAT and CART for shortening questionnaire lengths. Future research should consider administering varied assessments of patients over time to monitor their progress in multiple domains. For CR professionals, CART integrated with MyCARDSS would provide a feedback loop that informs the rehabilitation progress of their patients by providing real-time patient measurements.

(J Med Internet Res 2020;22(1):e12509) doi:10.2196/12509

KEYWORDS

psychometrics; computing methodologies; mHealth; internet; cardiac rehabilitation; needs assessment

Introduction

Background

Given the prominent role of the internet in many patients' lives nowadays, patient portals are increasingly deployed to involve patients in their care process. These portals, for example, allow fewer time-consuming consultations between patients and health care professionals by integrating batteries of questionnaires needed for diagnosis or as part of a patient's needs assessment. A precondition of this type of use of patient portals is the high quality of data to be exchanged. This is especially required when the patient portal is linked to a computerized decision support system (CDSS), used by health care providers for advice on therapy planning during the decision-making process. Extensive assessment procedures, however, may result in increased response burden on the patient, possibly resulting in low quality of response data [1-4].

There is a need for assessment procedures of shorter lengths that capture patient questionnaire data while attaining high data quality without undue response burden on patients. Computerized adaptive testing (CAT) methods, using Item Response Theory (IRT), have the potential to meet these needs and offer an attractive option to shorten questionnaires. A CAT algorithm uses information from questions already answered, to select the most appropriate question to be administered next. Therefore, a patient is offered only the fewest possible items. Chien et al [1] verified the effectiveness and efficacy of saving time and reducing the burden on patients through CAT applied on the Activities of Daily Living Scale. They found that mobile nursing services placed at the bedsides of patients could, through a CAT module, reduce the burden on patients and save time more than the traditional paper-and-pencil testing appraisals. Similarly, CAT-based administration of surveys of patient perception substantially reduced patient burden without compromising the precision of measuring patients' perceptions of hospitalization [2]. Another promising method is classification and regression tree (CART) analysis, originating from clinical decision rules research, which is mostly used to classify patients into clinically important categories [5,6]. It can be used to shorten questionnaires by selecting predictor variables (questionnaire items) that allow different questions to be identified for patients with different levels of complaints of a disease. As an illustration, Lu et al [7] successfully applied CART methods in the development of brief screening tools based on questions from existing psychiatric diagnostic instruments. Potential advantages of both CAT and CART for clinical practice are efficient testing and a reduction in the test burden in patients and, consequently, less measurement error during testing. However, as far as we know, CAT and CART performances in terms of their yield in shortening questionnaire lengths and sensitivity and specificity levels have not been compared.

In this study, we examined if CAT or CART analysis could be used to shorten the questionnaires included in the needs assessment procedure (NAP) of cardiac rehabilitation (CR) patients. CR is a therapy to support patients with cardiac issues in recovering from a cardiac incident in order to improve their physical and physiological condition [8]. To offer a patient a tailored rehabilitation plan, every patient has to complete an NAP including 80-130 questionnaire items of which answer data are sent to a CDSS. We aimed to test which method, CAT or CART, was best suited to reduce the number of questionnaire items in multiple domains (eg, anxiety, depression, quality of life, and social support) used for the NAP, without the loss of data quality.

Cardiac Rehabilitation in the Netherlands: Case Study

We used the data collected in the Dutch multicenter CARDSS program [8]. Within this program, CR clinics use a CARDSS electronic patient record (EPR) with computerized decision support (CDS) based on the most recent version of the Dutch CR guidelines [9]. The CDS provides CR professionals with advice on a patient-tailored rehabilitation program based on an NAP. The Dutch guideline requires gathering of 80-130 data items regarding a patient's quality of life, work resumption, psychological and social functioning, and lifestyle. The patient-tailored rehabilitation program can comprise four possible group-based therapies: disease-specific education; exercise training; lifestyle modification; and relaxation and stress management training, supplemented by individual counseling (eg, by a psychologist, dietician, or social worker) and, if needed, different forms of individual therapies. During the NAP, a CR professional can immediately discuss the CDS advice with the patient to set the final patient-tailored rehabilitation plan. To improve the efficiency of the NAP data gathering process, we developed an electronic portal for patients, called MyCARDSS, that patients can use to enter their data, either at home or at the CR clinic. Some patients fill in the NAP in MyCARDSS just before their consultation at the CR clinic, as they need the help of a nurse or do not have an internet connection at home [10]. MyCARDSS is linked to the EPR system.

Data of 2837 CR patients of this multicenter CARDSS Web-based program was included in this study. The CARDSS Web-based dataset comprised patient identification data and CR needs assessment data. We used the database of CARDSS Web-based program to obtain the scores on the individual questions of the patients' NAP data and the following demographic patient information: age, gender, diagnosis, and cardiac intervention.

Questionnaires Data

We included data from seven questionnaires used in the NAP for CR, which allow the classification of patients based on the outcome: no (low), mild (moderate), and serious (serious or high) symptoms. The following questionnaires are used in the multicenter CARDSS Web-based program:

The Dutch version of the Quality of Life after Myocardial Infarction (QLMI) is a 27-item questionnaire, scored on a 7-point Likert scale, to measure health-related quality of life for patients after myocardial infarction. It comprises 10 physical dimension items (QLMI-P), subscores ranging from 1 to 7, where subscale scores between 1.0 and 3.39 classify a patient as having a low and >4.0 as having a high exercise capacity; 7 social dimension items (QLMI-S), subscores ranging from 1 to

7, where subscale scores between 1.0 and 4.4 indicate a high, between 4.5 and 5.9 indicate a moderate, and >6.0 indicate a low risk on social dysfunctioning; and 10 emotional dimension items (QLMI-E) [11]. The QLMI-E is not used in the NAP for CR.

The Hospital Anxiety and Depression Scale (HADS) is a 14-item questionnaire, scored on a 4-point Likert scale (0-3), to detect the presence of anxiety and depression. It comprises seven anxiety items (HADS-A) and seven depression items (HADS-D). A total score is not calculated. For both subscales, scores range from 1 to 21. For both HADS-A and HADS-D, subscale scores between 0 and 4 points indicate a low, between 5 and 7 points indicate a moderate, and >8 points indicate a serious risk on anxiety [12].

The Patient Health Questionnaire–9 (PHQ-9) is a 9-item questionnaire on a 4-point Likert scale to detect the presence and severity of mental health disorders with scores ranging from 0 to 27: Scores between 0 and 4 points indicate a low, between 5 and 9 points indicate a moderate, and >10 points indicate a serious risk on depression [13].

The Generalized Anxiety Disorder-7 (GAD-7) is a 7-item questionnaire on a 4-point Likert scale to detect the presence of generalized anxiety disorders. Scores range from 0 to 21: Scores between 0 and 4 points indicate a low, between 5 and 9 points indicate a moderate, and >10 points indicate a serious risk on anxiety [14].

The Multidimensional Perceived Social Support Scale (MPSSS) is a 12-item questionnaire, scored on a 7-point Likert scale, to measure social support and specific availability and satisfaction with support from family, friends, or a special person. Scores range from 12 to 84: Scores between 12 and 64 points indicate a low, between 65 and 78 points indicate a moderate, and >79 points indicate a high level of social support [15].

The Dutch guidelines state that all patients have to fill in the QLMI-P/QLMI-S questionnaire, but CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in their patients. This means that an individual patient fills in the QLMI-P/QLMI-S and HADS-A/HADS-D or QLMI-P/QLMI-S and GAD-7/PHQ-9 questionnaires. Administration of the MPSSS is not mandatory; this questionnaire is used by merely one clinic in the CARDSS Web-based program to assess social support provided to a patient.

Methods

Classification and Regression Tree

We first used CART as a method to reduce the number of items of each questionnaire included in the NAP. CART represents a hierarchical model structured as trees for predicting continuous (regression trees) and categorical (classification trees) variables [16,17].

CART models predict a response variable (ie, outcome) based on the values of ≥1 predictors (ie, items within the questionnaire or demographic data) using a technique called recursive partitioning groups [18]. This algorithm looks for subgroups in the dataset in which the response variable is relatively homogeneous. These subgroups become the leaves of the tree [16,18]. At each node in the tree, the recursive partitioning algorithm identifies a predictor variable and a *split* by which cases may be subclassified. This predictor variable and split combination are chosen to have the greatest predictive power among all predictor split combinations at the tree node. Once the cases at the node have been partitioned by the split, the algorithm is applied to both resulting subclassifications. The bottom node of the tree reports a classification for the patient. In most cases, the constructed model is asymmetric, which means that it depends on previous answers on items administered to a patient. For qualitative outcome variables, the resulting tree is a classification tree, and for quantitative outcome variables, it is a regression tree. To shorten the NAP overall, all individual questions of each questionnaire were entered into a CART analysis to develop a CART per questionnaire. The CART analysis selects independent items that differentiate the outcome variable but allows different combinations of the predictor variables in different subgroups, creating flexible questionnaires [16,18]. CART allows the set of questionnaire items presented to a patient to be adapted to the responses already provided by him or her; going left at a node may result in a very different set of questionnaire items being presented as compared with going right. Thus, CART has the potential to shorten the length of the NAP. To determine if CART can shorten the NAP, we determined the mean, maximum, and minimum length of each CART per questionnaire included in the NAP.

We also tested if demographic or clinical variables—patient age, gender, cardiac diagnosis, and intervention—as predictors influenced the length of the CART.

Computer Adaptive Testing

The second method we used to shorten the number of questionnaire items of each questionnaire included in the NAP is CAT. The net result of a CAT is a small, optimal number of items to be administered to the patient without loss of measurement precision. CAT is based on IRT. IRT models are statistical models of the relationship between a person's score on the construct being measured and the probability of choosing each response on each item measuring that construct. IRT models can be used to evaluate how informative an item is for a specific range of scores and estimate a person's IRT score [19]. An IRT model expresses a probability (vertical axis) of the selection of each item response category as a function of the score (horizontal axis) on the underlying latent trait (the measured construct, ie, anxiety or depression in this study). To estimate the latent trait, a great number of different IRT models can be used [19]. For questions with ordered response categories, the Graded Response Model (GRM) has been proposed [19]. As the included NAP questionnaires have ordered response options, we fitted a GRM to our data to measure the item parameters of all questionnaire items. A CAT begins with an initial global question; all patients answer the same first item. On the basis of the response to the first item, the score, CI, and latent trait are estimated using maximum likelihood information [19]. The algorithm selects further items based on the highest possible information for the current latent trait score. The latent trait is estimated after each item administration based on the

accumulated information combined with the information on the new response [19]. The adaptive testing stops when a stopping rule is met. In this study, CAT stops as soon as a patient can be classified in a low, moderate, or high class with a CI of 95%. A patient had to have answered at least one item before the stopping rule was checked. With this stopping rule, the complete test cycle has a variable length, depending on the patient's individual responses and the point at which the stopping rule is applied. We did not set a minimum for item administration. Thus, theoretically, the algorithm could stop after the administration of only two or three items, given that the item is informative enough to classify a patient in the low, moderate, or high class. To determine if CAT can shorten the NAP, we determined the mean, maximum, and minimum length of each CAT per questionnaire included in the NAP.

Performance Testing

To determine the performance for both CAT and CART, the sensitivity and specificity were computed. For the regression trees, the root mean squared error (RMSE) and the normalized RMSE (NRMSE) were additionally measured. The RMSE is a measure of the differences (ie, the prediction errors) between values predicted by a regression/classification tree and those actually observed [18]. The RMSE serves to aggregate the prediction errors into a single measure of predictive power. The lower the RMSE or NRMSE, the better the model predicts.

Using the same data for calibration and evaluation of the model results in overly optimistic estimates of performance. Cross-validation is a method for validation of a procedure for model building that avoids the requirement for a new or independent validation set [20]. We, therefore, randomly split the data into two sets: a training set (2127/2837, 74.97%) for calibration of the models and a validation set (710/2837, 25.02%) for evaluating the performance of the CAT and CART. Besides, 10-fold cross-validation was divided into 10 subsets, each subset, in turn, being used to test the performance of the CAT/CART created with the other 9 subsets.

Software

All analyses were performed in R (R Foundation of Statistical Computing, Vienna, Austria), a programming language for statistical computing. Different packages were used to perform the analyses. R packages for simulating Item Response Theory based on Computerized Adaptive Tests and Latent Trait Modes packages were used for the CAT simulation. R packages for Classification and Regression Training and Recursive Partitioning And Regression Trees were used for CART. The training and test datasets were created with the *caret* package [21,22].

Results

Patient Characteristics

Demographic and clinical characteristics of the patients from the clinics participating in the CARDSS Web-based program are shown in Table 1.

Table 1. Demographic and clinical characteristics of 2837 cardiac rehabilitation patients.

Characteristics	Value
Gender, n (%)	
Male	1984 (69.99)
Female	826 (29.11)
Missing	27 (0.95)
Age (years), mean (SD)	
Men	65.8 (10.9)
Women	68.6 (11.3)
Mean	66.6 (11.1)
Diagnosis and intervention, n (%)	
ACS[a] (myocardial infarction or unstable angina pectoris) with intervention (CABG[b], PCI[c], CABGVALVE[d], or VALVESUR[e])	831 (29.29)
Chronic diagnosis (heart failure or stable angina pectoris)	648 (22.84)
Elective PCI (PCI without ACS)	604 (21.29)
Elective CABG (CABG, CABGVALVE, or VALVESUR without ACS)	404 (14.24)
ACS without intervention	350 (12.33)

[a]ACS: acute coronary syndrome.

[b]CABG: coronary artery bypass grafting.

[c]PCI: percutaneous coronary intervention.

[d]CABGVALVE: coronary artery bypass grafting in combination with heart valve surgery.

[e]VALVESUR: heart valve surgery.

Of the 2837 patients who participated in the program. 69.99% (1984/2837) were male and 29.29% (831/2837) had an acute coronary syndrome; men were younger (mean age 65.8, SD 10.9 years) than women (mean age 68.6, SD 11.3 years).

Inclusion and Exclusion of Questionnaires

Table 2 provides an overview of the total number of questionnaires fully filled out, missing questionnaires, and insufficiently filled out questionnaires. As explained previously, the Dutch guidelines state that CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in their patients. Administration of the MPSSS is not mandatory; this questionnaire is used by merely one clinic in the CARDSS Web-based program. Patient data on questionnaires were excluded from the CAT and CART analysis if (1) a questionnaire was not filled out by the patient (missing) or (2) a provided questionnaire was insufficiently filled out by a patient to calculate a total score (≥1 item responses missing).

Table 2. Number of patients who filled out, did not fill out, or insufficiently filled out the questionnaires (N=2837).

Questionnaire	Fully filled out, n (%)	Missing, n (%)	Insufficiently filled out, n (%)
Quality of Life after Myocardial Infarction - Physical dimension	2633 (92.81)	189 (6.66)	15 (0.52)
Quality of Life after Myocardial Infarction - Social dimension	2633 (92.81)	189 (6.66)	15 (0.52)
Patient Health Questionnaire - 9	1156 (40.75)	213 (7.50)	0 (0.00)
Generalized Anxiety Disorder - 7	1223 (43.11)	244 (8.60)	1 (0.04)
Hospital Anxiety and Depression Scale - Anxiety	1266 (44.62)	152 (5.35)	2 (0.07)
Hospital Anxiety and Depression Scale - Depression	1264 (44.55)	149 (5.25)	4 (0.14)
Multidimensional Perceived Social Support Scale	716 (25.23)	57 (2.00)	5 (0.17)

Classification and Regression Tree

CART models were built for every questionnaire in the NAP using the training set, resulting in a total of 28 CART models: (1) seven classification trees without the additional clinical/demographic data (ie, age, gender, cardiac diagnosis, and intervention) as features, (2) seven classification trees with the additional clinical/demographic data as features, (3) seven regression trees without the additional clinical/demographic data as features, and (4) seven regression trees with the additional clinical/demographic data as features.

The seven classification and seven regression trees with data on the clinical/demographic variables—age, gender, diagnosis, and intervention—showed no inclusion of these variables in the classification or regression trees.

The developed classification trees comprise four to six levels, with five to eight terminal nodes. The developed regression trees comprise four levels, with five terminal nodes.

Performances of Computerized Adaptive Testing and Classification and Regression Tree

To evaluate the performances of CAT and CART, patient data in the validation dataset were used. Figure 1 shows the maximum, minimum, and mean number of items administered by CART and CAT. Figure 2 displays the percentage decrease per questionnaire for CART and CAT. Table 3 lists the performances of both CAT and CART in terms of sensitivity and specificity per questionnaire. These performance measures are provided for each of the categories—low, moderate, and high—except for the QLMI-P with only low and high classes.

Figure 1. Maximum, minimum, and mean number of items administered in each CAT (computerized adaptive testing) and classification and regression tree. GAD-7: Generalized Anxiety Disorder - 7; HADS-A: Hospital Anxiety and Depression Scale - Anxiety; HADS-D: Hospital Anxiety and Depression Scale - Depression; MPSSS: Multidimensional Perceived Social Support Scale; PHQ-9: Patient Health Questionnaire - 9; QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension; and QLMI-S: Quality of Life after Myocardial Infarction - Social dimension.

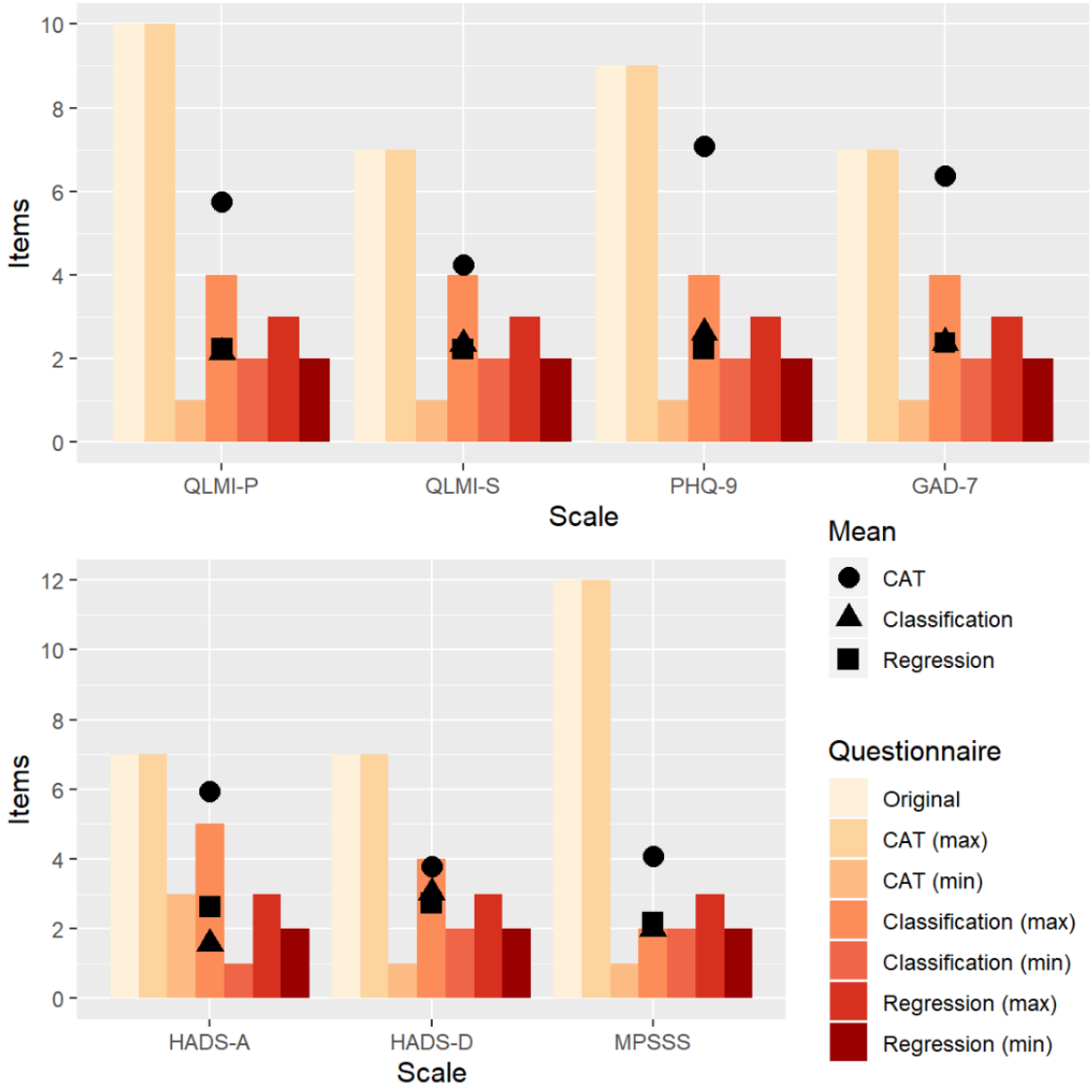


Figure 2. Percentage decrease per questionnaire for computerized adaptive testing and classification and regression tree. QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension; QLMI-S: Quality of Life after Myocardial Infarction - Social dimension; PHQ-9: Patient Health Questionnaire - 9; GAD-7: Generalized Anxiety Disorder - 7; HADS-A: Hospital Anxiety and Depression Scale - Anxiety; HADS-D: Hospital Anxiety and Depression Scale -Depression; MPSSS: Multidimensional Perceived Social Support Scale.

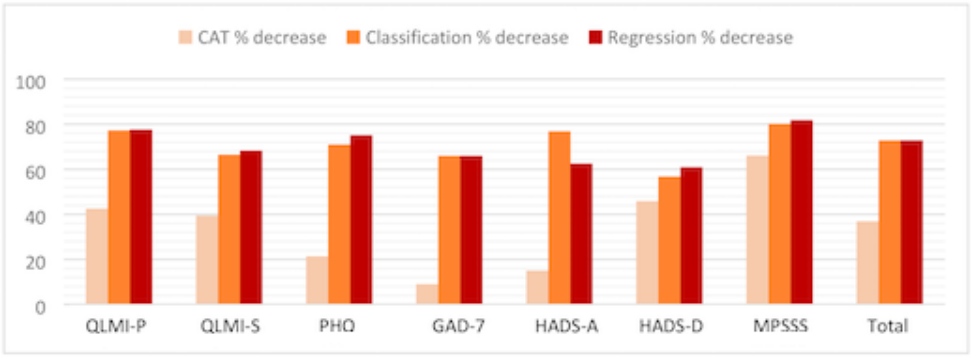


Table 3. Sensitivity and specificity of computerized adaptive testing, classification trees, and regression trees by questionnaire (sensitivity and specificity of a class are computed as the class versus the other classes, ie, low versus moderate and high/serious, moderate versus low and high/serious, and high/serious versus low and moderate). High scores on Patient Health Questionnaire - 9, Generalized Anxiety Disorder - 7, Hospital Anxiety and Depression Scales but conversely low scores on Multidimensional Perceived Social Support Scale and Quality of Life after Myocardial Infarction - Physical and Social dimensions instrument indicate that patients report surveyed symptoms.

Questionnaire	Sensitivity			Specificity			Root mean squared error	Normalized root mean squared error
	Low	Moderate	High	Low	Moderate	High		
Computerized adaptive testing								
QLMI-P[a]	0.609[b]	N/A[c]	N/A	1[b]	N/A	N/A	N/A	N/A
QLMI-S[d]	1	0.008	0.201	0.339	0.691	1	N/A	N/A
PHQ-9[e]	0.773	0.000	0.988[f]	0.887	0.997	.0581[f]	N/A	N/A
GAD-7[g]	0.935	0.222	1[f]	0.766	0.996	0.852[f]	N/A	N/A
HADS-A[h]	0.765	0.333	0.884[f]	0.961	0.887	0.810[f]	N/A	N/A
HADS-D[i]	0.234	0	1[f]	1	1	0.126[f]	N/A	N/A
MPSS[j]	1	0	0.515	0.256	1	1	N/A	N/A
Classification trees								
QLMI-P	0.954[b]	N/A	N/A	0.763[b]	N/A	N/A	N/A	N/A
QLMI-S	0.734	0.735	0.807	0.947	0.786	0.892	N/A	N/A
PHQ-9	0.907	0.554	0.549[f]	0.852	0.843	0.922[f]	N/A	N/A
GAD-7	0.931	0.681	0.800[f]	0.869	0.921	0.964[f]	N/A	N/A
HADS-A	0.941	0.417	0.970[f]	0.969	0.970	0.879[f]	N/A	N/A
HADS-D	0.979	0.275	0.936[f]	0.938	0.960	0.793[f]	N/A	N/A
MPSSS	0.644	0.836	0.758	0.970	0.712	0.938	N/A	N/A
Regression trees								
QLMI-P	0.946[b]	N/A	N/A	0.640[b]	N/A	N/A	0.661	0.247
QLMI-S	0.849	0.648	0.755	0.919	0.787	0.871	0.608	0.184
PHQ-9	0.861	0.551	0.832[f]	0.833	0.857	0.945[f]	2.696	0.222
GAD-7	0.957	0.408	0.694[f]	0.645	0.933	0.977[f]	2.231	0.145
HADS-A	0.750	0.667	0.922[f]	0.966	0.907	0.966[f]	1.833	0.153
HADS-D	0.640	0.190	0.985[f]	0.996	0.917	0.667[f]	1.582	0.146
MPSSS	0.489	0.647	0.923	0.977	0.755	0.805	7.962	0.162

[a]QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension.

[b]The quality of Life after Myocardial Infarction - Physical dimension has only a low and high class, so the sensitivity and specificity are computed as low versus high.

[c]Not applicable.

[d]QLMI-S: Quality of Life after Myocardial Infarction - Social dimension.

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

[f]Actual class is serious.

[g]GAD-7: Generalized Anxiety Disorder - 7.

[h]HADS-A: Hospital Anxiety and Depression Scale - Anxiety.

[i]HADS-D: Hospital Anxiety and Depression Scale - Depression.

[j]MPSSS: Multidimensional Perceived Social Support Scale.

Classification and Regression Tree

The minimum and maximum number of questionnaire items to be selected for the different questionnaires, based on the training set, are shown in Figure 1. For the classification trees, an average of 2.4 items per questionnaire should be administered (minimum=2 and maximum=3) to classify a patient (Figure 1), with a questionnaire reduction between 60.9% (HADS-D) and 77.6% (QLMI-P), compared with the original questionnaire (Figure 2). For the regression trees, an average of 2.4 items per questionnaire should be administered (minimum=1 and maximum=5) to classify a patient (Figure 1), with a questionnaire reduction between 56.8% (HADS-D) and 77.4% (QLMI-P), compared with the original questionnaire (Figure 2).

For all questionnaires except the HADS-A, the minimum number of questionnaire items (two items) in the classification trees equals those of the regression trees. The maximum number of questionnaire items is higher in the classification (four or five items) than in the regression (three items for all questionnaires) trees, except for the MPSSS (two items) and QLMI-P (three items). The mean number of items and the mean percentage decrease over all questionnaires in the classification trees equal those of the regression trees (2.38). A mean 72% reduction over all questionnaires in the NAP can be realized by CART.

Per questionnaire, the sensitivity and specificity levels of CARTs, classification, and regression trees are displayed in Table 3. The mean sensitivity and specificity of CART over all questionnaires and classes are 0.777 and 0.877, respectively. The mean sensitivity and specificity of regression trees over all questionnaires and classes are 0.743 and 0.840, respectively. The NRMSE of the regression trees ranges from 0.145 (GAD-7) to 0.247 (QLMI-P).

Computerized Adaptive Testing

The minimum and maximum number of questionnaire items used by CAT to be selected for the different questionnaires, based on the training set, are shown in Figure 1. With CAT, an average of 5.3 items (minimum=1 and maximum=original length) were needed per questionnaire to classify a patient (our goal was not to classify but to shorten the questionnaire), with a questionnaire reduction between 9.0% (GAD-7) and 45.8% (HADS-D), compared with the original questionnaire (Figure 2). CAT shows a smaller percentage decrease in questionnaire items per questionnaire and a smaller overall decrease, compared with CART (Figure 2). A mean 36% reduction over all questionnaires in the NAP can be realized by CAT. The sensitivity and specificity levels of CATs per questionnaire are displayed in Table 3. The mean sensitivity and specificity of CAT over all questionnaires and classes are 0.765 and 0.817, respectively.

Computerized Adaptive Testing and Classification and Regression Tree: Questionnaire Comparison

The differences in percentage decrease in questionnaire items per questionnaire between CAT and CART are highest for the PHQ-9 (56.7%), GAD-7 (56.5%), and HADS-A (54.6%). HADS-D (13.0%) and MPSSS (14.9%) show the lowest differences in percentage decrease per questionnaire (Figure 2).

Discussion

Principal Findings

This study shows that both CAT and CART can be used to shorten the questionnaires of the NAP used within the field of CR. CART, however, showed the best performance with an overall about twice as large decrease in questionnaire items of the NAP and the highest sensitivity and specificity. Demographic/clinical variables—patient age, gender, cardiac diagnosis, and intervention—as predictors did not influence the length of the CARTs, meaning that these variables do not determine the classification of patients in the trees.

Relation to Other Studies

CAT has nearly four decades of research behind it but has only been applied more recently to health care. CAT has been used to shorten or develop questionnaires for assessment of fatigue [23], depression [24-26], suicide ideation [4], other mental health disorders [27,28], physical [29] and upper extremity functioning [30], health status in patients with knee osteoarthritis [31], activities of daily living in outpatients with stroke [32], and exposure of nurses to workplace bullying [33,34] and in patient-reported outcome measurement studies [24,29]. Overall, its application has proven to be successful in shortening questionnaires, while patient measurements remained valid and reliable. Some studies even demonstrated that by applying CAT, existing instruments for patient-reported outcomes could be improved. These new instruments reduced the questionnaire burden on patients while increasing measurement precision [29], possibly leading to reduced sample size requirements.

Similarly, half a century has passed since the publication of the first CART algorithms, but again, their application in health care is of a far more recent date.

CART has, for the most part, been used to classify (new) patients into clinically important (risk) categories, such as diabetic nephropathy [35] and colorectal adenocarcinoma [36]. CART has also been applied to define factors associated with delayed treatment of acute myocardial infarction [37] and quality of life [38].

As far as we know, CART has not been used, at least not in the health care domain, with the aim to shorten questionnaires.

To our knowledge, our study is the first to assess the differences in performance between CAT and CART for shortening questionnaire lengths. Overall, CART outperformed CAT, with a larger reduction in the length of the questionnaire for the NAP procedure and in sensitivity and specificity.

Further, we did not observe an influence of the predictors such as age, gender, diagnosis, and intervention in the construction of CARTs. CART would have probably captured interactions across many NAP questionnaire scores and these clinical/demographic variables. Our findings are in concordance with the findings of the study by Miscio et al [39] wherein the inclusion of clinical/demographic data such as age and gender

did not have an effect on the construction of CARTs for patient measurement tools.

Meaning of This Study

Ideally, an NAP procedure for CR including several questionnaires should be highly sensitive and specific, so that few patients with depression/anxiety/social complaints are missed and few without depression/anxiety/social complaints are identified as having complaints. In this context, CART is, overall, more sensitive and specific than CAT, while it shortens the NAP procedure more than CAT. CART analysis has the statistical advantage of being a nonparametric method, with no assumptions about the functional form of the data. CART might further be a good alternative to CAT, as this method not only has the ability to efficiently shorten questionnaires by segmenting patient groups into meaningful subgroups, but it also presents knowledge on these subgroups in a graphical way. These graphs provide a good understanding of how this segmentation was attained. CART, as an algorithmic rather than statistical method, further offers good insight into interactions between variables that are not revealed by linear quantitative research. But CART does not provide distributions, likelihood ratios, or CIs to quantify or support the validity of the findings. For the CATs, we, for example, made use of CIs; we stopped a CAT as soon as a patient could be classified in a low, moderate, or high class with a CI of 95%. For evaluating the performance of both the CAT and CART, we made use of cross-validation by splitting the dataset in a training set for calibration of the models and a validation set. We further applied 10-fold cross-validation on the training and validation sets to validate the generality of the CAT and CARTs.

Recommendations for Care Practice and Future Research

We demonstrated that an NAP with shortened questionnaires can be used for screening cardiac patients on rehabilitation needs without compromising its measurement accuracy. Both CAT and CART can be used for this purpose, but a CART approach with many questionnaires is rare. Obtaining the information during the joint administration of the shortened questionnaires would take about one-third of the time that it would take with the traditional fixed-length questionnaires, that is, 30 min instead of 90 min per patient. We also plan to extend the MyCARDSS portal with adaptive tests that will be administered to our CR patients over time, at their home or just before their consultation at the clinic. The results of these assessments will be interfaced with the CARDSS EPR, which is easily accessible by the CR professionals from any device. For CR professionals, CART integrated with MyCARDSS and CARDSS would then provide a feedback loop that informs the rehabilitation progress of their patients by providing these real-time patient measurements. An example for future research is to monitor our CR patients' physical and mental health progress by these varied assessments. Although dimension reduction strategies have been employed for numerous data problems, they are scantly discussed in the context of analyzing survey data. Another example for future research, thus, is to examine the performance of other methods such as dimensionality reduction techniques, for example, principal component analyses (PCA) for reducing questionnaire lengths. Dimension reduction techniques can be an effective approach for reducing dimensionality in more complex survey data sources than the data source used in this study. Methods such as CART, CAT, and PCA could, for example, be applied on federal and other publicly available datasets to further improve the validity and generalizability of the findings of this case study and conduct more efficient and cost-effective surveys in the future.

Strengths and Limitations

A strength of this study pertains to its large sample size; 2837 patients from various CR clinics completed the various NAP questionnaires, confirming the external validity of the findings. The number of fully filled out questionnaires per type of questionnaire varied from 716 to 2633, with MPSSS being used in one CR clinic and having the lowest response rate. The original version of the MPSSS, furthermore, has a low number of questionnaire items in comparison with QLMI, for example. This might have impacted the CAT and CART analysis. It, hence, remains uncertain whether the results, particularly those of MPSSS, can be generalized to other CR clinics. An even larger sample size could have led to a more precise estimation of NAP questionnaire lengths needed to determine patients' specific CR needs.

Another strength of this study is the application, exploration, and comparison of the two techniques for shortening questionnaires. The use of CART analysis as one of these techniques has been suboptimal at the least; we did not find any study using CART aiming at reducing questionnaire lengths. Finally, the Dutch guideline for CR prescribes that CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in patients. We do not know if these combinations of questionnaires are equally valid in measuring the constructs of anxiety and depression. With CART, a combination of HADS-A/HADS-D would result in a reduction in questionnaire lengths of six and five items, respectively, whereas the combined GAD-7/PHQ-9 would lead to a reduction of seven and five items, respectively.

Finally, we did not examine if the questionnaires of the NAP produced similar results when patients filled them in on the Web through MyCARDSS at their home or in the CR clinic.

Conclusions

CART and CAT both have shown to be accurate methods for reducing the length of the NAP used in the field of CR. Of both methods, CART overall showed the largest decrease in the number of questionnaire items and the best performance in terms of sensitivity and specificity. This study is the first to apply and compare the performances of CAT and CART for shortening questionnaires, and it demonstrated that the use of CART analysis would be a step forward in the development of a shorter NAP questionnaire for CR patients and possibly other questionnaires.

Instead of using long, fixed-length questionnaires on paper or on the Web for the NAP, a much smaller set of questionnaire items will suffice to identify patients' varying needs for CR,

without the loss of information and data quality and an excessive burden on the patient.

Acknowledgments

We would like to thank A Mulder-Wiggers for her assistance in the literature search for this study.

Conflicts of Interest

None declared.

References

1. Chien T, Wu H, Wang W, Castillo R, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. Health Qual Life Outcomes 2009 May 5;7:39 [FREE Full text] [doi: 10.1186/1477-7525-7-39] [Medline: 19416521]
2. Chien T, Wang W, Huang S, Lai W, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. J Med Internet Res 2011 Aug 15;13(3):e61 [FREE Full text] [doi: 10.2196/jmir.1785] [Medline: 21844001]
3. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]
4. de Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof A. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. J Med Internet Res 2014 Sep 11;16(9):e207 [FREE Full text] [doi: 10.2196/jmir.3511] [Medline: 25213259]
5. Zimmerman R, Balasubramani G, Nowalk M, Eng H, Urbanski L, Jackson M, et al. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infect Dis 2016 Sep 22;16(1):503 [FREE Full text] [doi: 10.1186/s12879-016-1839-x] [Medline: 27659721]
6. Valera V, Walter B, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 2007 Jan;14(1):34-40. [doi: 10.1245/s10434-006-9145-2] [Medline: 17024555]
7. Lu F, Petkova E. A comparative study of variable selection methods in the context of developing psychiatric screening instruments. Stat Med 2014 Feb 10;33(3):401-421 [FREE Full text] [doi: 10.1002/sim.5937] [Medline: 23934941]
8. van Engen-Verheul MM, de Keizer NF, van der Veer SN, Kemps HM, Reimer WJ, Jaspers MW, et al. Evaluating the effect of a web-based quality improvement system with feedback and outreach visits on guideline concordance in the field of cardiac rehabilitation: rationale and study protocol. Implement Sci 2014 Dec 31;9:780 [FREE Full text] [doi: 10.1186/s13012-014-0131-y] [Medline: 25551791]
9. Rehabilitation CNSFC(HF(GPAAHR. Multidisciplinary Guidelines for Cardiac Rehabilitation (in Dutch). Utrecht: Netherlands Society for Cardiology (NVVC) 2012:1-16.
10. Wiggers A, Vosbergen S, Kraaijenhagen R, Jaspers M, Peek N. Changes in the cardiac rehabilitation workflow process needed for the implementation of a self-management system. Stud Health Technol Inform 2013;192:1140. [doi: 10.3233/978-1-61499-289-9-1140] [Medline: 23920914]
11. Dempster M, Donnelly M, O'Loughlin C. The validity of the MacNew Quality of Life in heart disease questionnaire. Health Qual Life Outcomes 2004 Jan 22;2:6 [FREE Full text] [doi: 10.1186/1477-7525-2-6] [Medline: 14738566]
12. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. J Psychosom Res 1997 Jan;42(1):17-41. [doi: 10.1016/s0022-3999(96)00216-4] [Medline: 9055211]
13. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
14. Spitzer R, Kroenke K, Williams J, 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]
15. Dahlem N, Zimet G, Walker R. The Multidimensional Scale of Perceived Social Support: a confirmation study. J Clin Psychol 1991 Nov;47(6):756-761. [doi: 10.1002/1097-4679(199111)47:6<756::aid-jclp2270470605>3.0.co;2-l] [Medline: 1757578]
16. Alpaydin E. Introduction to Machine Learning. Second Edition. London: MIT Press; 2010.
17. Song Y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015 Apr 25;27(2):130-135 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.215044] [Medline: 26120265]
18. Breiman L, Friedman J, Stone C, Olshen R. Classification And Regression Trees. London: Chapman And Hall; 2017.
19. van der Linden WJ, Glas CA. Elements of Adaptive Testing. An Advanced and Comprehensive Introduction to the Foundation of CAT. Amsterdam: Springer; 2010.
20. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv 2010;4:40-79. [doi: 10.1214/09-SS054]

21. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
22. Therneau TM, Atkinson EJ. CRAN - R Project. 2018. An Introduction to Recursive Partitioning Using the RPART Routines URL: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf [accessed 2019-09-15]
23. Nikolaus S, Bode C, Taal E, Vonkeman HE, Glas CA, van de Laar MA. Construct validation of a multidimensional computerized adaptive test for fatigue in rheumatoid arthritis. PLoS One 2015;10(12):e0145008 [FREE Full text] [doi: 10.1371/journal.pone.0145008] [Medline: 26710104]
24. Flens G, Smits N, Terwee C, Dekker J, Huijbrechts I, de Beurs E. Development of a computer adaptive test for depression based on the Dutch-Flemish version of the PROMIS item bank. Eval Health Prof 2017 Mar;40(1):79-105. [doi: 10.1177/0163278716684168] [Medline: 28705028]
25. Gibbons R, Weiss D, Pilkonis P, Frank E, Moore T, Kim J, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012 Nov;69(11):1104-1112 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.14] [Medline: 23117634]
26. Gibbons R, Hooker G, Finkelman M, Weiss D, Pilkonis P, Frank E, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. J Clin Psychiatry 2013 Jul;74(7):669-674 [FREE Full text] [doi: 10.4088/JCP.12m08338] [Medline: 23945443]
27. Choi SW, van der Linden WJ. Ensuring content validity of patient-reported outcomes: a shadow-test approach to their adaptive measurement. Qual Life Res 2018 Jul;27(7):1683-1693. [doi: 10.1007/s11136-017-1650-1] [Medline: 28710673]
28. Gibbons R, Weiss D, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016;12:83-104. [doi: 10.1146/annurev-clinpsy-021815-093634] [Medline: 26651865]
29. Fries J, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014 Jan;41(1):153-158. [doi: 10.3899/jrheum.130813] [Medline: 24241485]
30. Lehman L, Woodbury M, Shechtman O, Wang Y, Pomeranz J, Gray D, et al. Development of an item bank for a computerised adaptive test of upper-extremity function. Disabil Rehabil 2011;33(21-22):2092-2104. [doi: 10.3109/09638288.2011.560336] [Medline: 21401332]
31. Öztuna D, Elhan A, Küçükdeveci AA, Kutlay S, Tennant A. An application of computerised adaptive testing for measuring health status in patients with knee osteoarthritis. Disabil Rehabil 2010;32(23):1928-1938. [doi: 10.3109/09638281003777572] [Medline: 20384449]
32. Hsueh I, Chen J, Wang C, Hou WH, Hsieh CL. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. Phys Ther 2013 May;93(5):681-693. [doi: 10.2522/ptj.20120173] [Medline: 23329557]
33. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]
34. Ma S, Wang H, Chien T. A new technique to measure online bullying: online computerized adaptive testing. Ann Gen Psychiatry 2017;16:26 [FREE Full text] [doi: 10.1186/s12991-017-0149-z] [Medline: 28680455]
35. Sayyad M, Gopal G, Shahani A. Classification and regression trees: a possible method for creating risk groups for progression to diabetic nephropathy. J Appl Sci 2011;11(12):2076-2083. [doi: 10.3923/jas.2011.2076.2083]
36. Chang YJ, Chen LJ, Chang YJ, Chung KP, Lai MS. Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. BMC Med Res Methodol 2012 Jan 3;12:2 [FREE Full text] [doi: 10.1186/1471-2288-12-2] [Medline: 22214198]
37. Kuhn L, Worrall-Carter L, Ward J, Page K. Factors associated with delayed treatment onset for acute myocardial infarction in Victorian emergency departments: a regression tree analysis. Australas Emerg Nurs J 2013 Nov;16(4):160-169. [doi: 10.1016/j.aenj.2013.08.002] [Medline: 24199901]
38. Su X, Azuero A, Cho J, Kvale E, Meneses K, McNees M. An introduction to tree-structured modeling with application to quality of life data. Nurs Res 2011;60(4):247-255 [FREE Full text] [doi: 10.1097/NNR.0b013e318221f9bc] [Medline: 21720217]
39. D'Alisa MS, Miscio G, Baudo S, Simone A, Tesio L, Mauro A. Depression is the main determinant of quality of life in multiple sclerosis: a classification-regression (CART) study. Disabil Rehabil 2006 Mar 15;28(5):307-314. [doi: 10.1080/09638280500191753] [Medline: 16492625]

Abbreviations

CART: classification and regression tree
CAT: computerized adaptive testing
CDS: computerized decision support
CDSS: computerized decision support system
CR: cardiac rehabilitation
EPR: electronic patient record
GAD-7: Generalized Anxiety Disorder - 7

GRM: Graded Response Model
HADS-A: Hospital Anxiety and Depression Scale - Anxiety
HADS-D: Hospital Anxiety and Depression Scale - Depression
IRT: Item Response Theory
MPSSS: Multidimensional Perceived Social Support
NAP: needs assessment procedure
NRMSE: normalized root mean squared error
PCA: principal component analyses
PHQ-9: Patient Health Questionnaire - 9
QLMI-E: Quality of Life after Myocardial Infarction - Emotional dimension
QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension
QLMI-S: Quality of Life after Myocardial Infarction - Social dimension
RMSE: root mean squared error

Edited by G Eysenbach; submitted 17.10.18; peer-reviewed by D Mendes, C Reis, M Boukhechba; comments to author 07.02.19; revised version received 11.06.19; accepted 19.07.19; published 30.01.20.

Please cite as:
Peute L, Scheeve T, Jaspers M
Classification and Regression Tree and Computer Adaptive Testing in Cardiac Rehabilitation: Instrument Validation Study
J Med Internet Res 2020;22(1):e12509
URL: https://www.jmir.org/2020/1/e12509
doi:10.2196/12509
PMID:32012065

Original Paper

PediTools Electronic Growth Chart Calculators: Applications in Clinical Care, Research, and Quality Improvement

Joseph H Chou[1,2], MD, PhD; Sergei Roumiantsev[1,2], MD, PhD; Rachana Singh[3,4], MD, MS

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

[2]Harvard Medical School, Boston, MA, United States

[3]Baystate Children's Hospital, Springfield, MA, United States

[4]University of Massachusetts Medical School, Springfield, MA, United States

Corresponding Author:
Joseph H Chou, MD, PhD
Massachusetts General Hospital
55 Fruit Street, Founders 526E
Boston, MA, 02114-2696
United States
Phone: 1 6177249040
Fax: 1 6177249346
Email: jchou2@mgh.harvard.edu

Abstract

Background: Parameterization of pediatric growth charts allows precise quantitation of growth metrics that would be difficult or impossible with traditional paper charts. However, limited availability of growth chart calculators for use by clinicians and clinical researchers currently restricts broader application.

Objective: The aim of this study was to assess the deployment of electronic calculators for growth charts using the lambda-mu-sigma (LMS) parameterization method, with examples of their utilization for patient care delivery, clinical research, and quality improvement projects.

Methods: The publicly accessible PediTools website of clinical calculators was developed to allow LMS-based calculations on anthropometric measurements of individual patients. Similar calculations were applied in a retrospective study of a population of patients from 7 Massachusetts neonatal intensive care units (NICUs) to compare interhospital growth outcomes (change in weight Z-score from birth to discharge [ΔZ weight]) and their association with gestational age at birth. At 1 hospital, a bundle of quality improvement interventions targeting improved growth was implemented, and the outcomes were assessed prospectively via monitoring of ΔZ weight pre- and postintervention.

Results: The PediTools website was launched in January 2012, and as of June 2019, it received over 500,000 page views per month, with users from over 21 countries. A retrospective analysis of 7975 patients at 7 Massachusetts NICUs, born between 2006 and 2011, at 23 to 34 completed weeks gestation identified an overall ΔZ weight from birth to discharge of –0.81 (P<.001). However, the degree of ΔZ weight differed significantly by hospital, ranging from –0.56 to –1.05 (P<.001). Also identified was the association between inferior growth outcomes and lower gestational age at birth, as well as that the degree of association between ΔZ weight and gestation at birth also differed by hospital. At 1 hospital, implementing a bundle of interventions targeting growth resulted in a significant and sustained reduction in loss of weight Z-score from birth to discharge.

Conclusions: LMS-based anthropometric measurement calculation tools on a public website have been widely utilized. Application in a retrospective clinical study on a large dataset demonstrated inferior growth at lower gestational age and interhospital variation in growth outcomes. Change in weight Z-score has potential utility as an outcome measure for monitoring clinical quality improvement. We also announce the release of open-source computer code written in R to allow other clinicians and clinical researchers to easily perform similar analyses.

(J Med Internet Res 2020;22(1):e16204) doi:10.2196/16204

KEYWORDS

growth charts; pediatrics; infant, newborn; infant, premature; failure to thrive; internet; software

Introduction

Background

Failure to thrive secondary to inadequate nutrition in the pediatric population may result in lifelong negative impact on physical and mental health outcomes [1,2]. This is especially critical for infants and children with known risk factors, such as preterm birth, acute and chronic illnesses, and social risk factors [3-7].

Anthropometric measurements commonly used in pediatric populations to assess nutritional status include weight, length, stature, head circumference, and midarm circumference. Using appropriate growth chart references, a single measurement alone indicates growth status for age at a single time point and may provide indications for closer monitoring. With multiple measurements, growth velocity over time can be evaluated and deviation from normal growth pattern may be suggestive of suboptimal nutrition or chronic illnesses, including metabolic disorders or congenital syndromes, although suboptimal monitoring itself may impact efficacy [8].

Before more widespread availability of electronic health records, paper growth charts were commonly used, but they had limitations, including infrequent updating, restricted accessibility for multiple care providers, and the inability to exactly determine percentiles numerically between the limited discrete percentile lines displayed on the printed charts.

The development of the lambda-mu-sigma (LMS) method for describing growth charts allows a quantitative description of growth charts based on tables of parameters [9]. In these tables, parameters for anthropometric measurements of interest relate a measurement at a given age to a precisely calculated Z-score (number of SDs from the mean) and percentile. Similarly, the expected anthropometric measurement at a particular Z-score and age can also be calculated. The availability of the LMS method and parameters for an increasing number of growth charts provides an opportunity to both improve clinical care of individual patients and allow large-scale analysis of datasets, which would be difficult or impossible if using paper growth charts.

Postnatal growth failure is common in preterm infants and is known to be associated with long-term neurodevelopmental impairment [10-18]. Extending the calculation of anthropometric measurement Z-scores from individual patients to a large population of patients might yield insight into how populations of preterm infants grow during their birth hospitalization. Similarly, we hypothesized that assessing the efficacy of quality improvement initiatives targeting improved growth might benefit from an unambiguous quantitative metric based on anthropometric Z-scores.

Objectives

In this paper, we describe the deployment of the publicly accessible PediTools website, which implements a suite of calculators supporting LMS-based growth charts. We hypothesized that a simple metric to assess growth outcomes—the change in weight Z-score from birth to discharge (ΔZ weight)—might yield insight into growth outcome variations. We retrospectively compared outcomes at 7 Massachusetts neonatal intensive care units (NICUs) and further utilized this metric to assess the efficacy of a nutrition-based quality improvement project at one of the NICUs. In addition, we also announce the release of open-source software, which will allow others to perform large-scale LMS-based calculations more easily.

Methods

Lambda-Mu-Sigma Method of Describing Growth References

The LMS method allows a parametric definition of growth references and generation of smoothed centile curves accounting for skewness of the distribution of an anthropometric measurement [9]. The parameters lambda (L, skewness normalization via power in the Box-Cox transformation), mu (M, mean), and sigma (S, coefficient of variation) describe the distribution of the measurement (eg, weight, length, or head circumference) at a given age, and the set of LMS parameters across multiple ages parameterizes the entire growth chart. This allows convenient calculation of exact Z scores (SDs from the mean) and generation of any centile curve.

Obtaining Lambda-Mu-Sigma Parameters

LMS parameters for growth charts were obtained either from the original publications, Web-based electronic supplements to the publications, and internet archives or by licensing agreement with the publication authors (references and sources listed in Table 1).

Table 1. Anthropometric growth calculators implemented on PediTools and sources of lambda-mu-sigma parameters.

Chart	Age range	Measures
Fenton 2003 preterm [19,20]	22-50 weeks gestation	Weight, head circumference, and length
Fenton 2013 preterm [21]	22-50 weeks gestation	Weight, head circumference, and length
CDC[a] infant [22,23]	0-36 months	Weight, head circumference, and length
CDC pediatric [22,23]	24-240 months	Weight, height, and BMI
WHO[b] infant [24]	0-24 months	Weight, head circumference, and length
Olsen preterm [25]	23-41 weeks gestation	Weight, head circumference, and length
WHO [26]	3-60 months	Arm circumference and triceps and subscapular skinfolds
CDC [27]	2-20 years	triceps and subscapular skinfolds
Olsen preterm BMI [28]	24-41 weeks gestation	BMI
Down syndrome infant [29,30]	0-36 months	Weight, length, and head circumference
Down syndrome pediatric [29,30]	2-20 years	Weight, height, head circumference, and BMI
CDC arm circumference [31]	2-222 months	Arm circumference
Mramba arm circumference [32]	60-228 months	Arm circumference

[a]CDC: Centers for Disease Control and Prevention.

[b]WHO: World Health Organization.

Interpolation of Lambda-Mu-Sigma Values

For each growth chart described via the LMS method, the L, M, and S curves are smoothed over ages, which permits interpolation of appropriate LMS values for intermediate values among the available discrete ages. In the PediTools calculators, simple linear interpolation was performed to obtain LMS values for intermediate ages. Different charts provide different degrees of age granularity. The Centers for Disease Control and Prevention (CDC) infant charts provide LMS parameters for ages in 1-month intervals, centered at the half-month point for the entire month [22], whereas the Fenton 2003 preterm charts provide parameters for completed weeks of gestation, centered midweek, for example, 30 weeks of completed gestation is centered around 30 3/7 weeks [19,20]. In contrast, the LMS values obtained for the Fenton 2013 preterm charts have values defined for each day of gestation; therefore, interpolation is not required [21].

Calculations via the Lambda-Mu-Sigma Method

Calculations of a Z score from LMS parameters and a given anthropometric measurement or an anthropometric measurement at a given Z score and LMS parameters at a particular age were performed as previously described (Figure 1) [9,22]. In the PediTools Web-based calculators, the percentile corresponding to a Z score was calculated by a numerical estimation of the cumulative density function of the standard normal distribution (equation 26.2.17 in the reference by Abramowitz et al) [33]. For the peditools R package, the same functionality is provided in the standard R function pnorm().

Figure 1. Equations for LMS-based growth metric calculations for Z score (a and b) and for an anthropometric measurement X (c and d).

a) $$Z = \frac{(\frac{X}{M})^{L} - 1}{LS} \quad \text{for } L \neq 0$$

b) $$Z = \frac{\ln(\frac{X}{M})}{S} \quad \text{for } L = 0$$

c) $$X = M(1 + LSZ)^{\frac{1}{L}} \quad \text{for } L \neq 0$$

d) $$X = Me^{SZ} \quad \text{for } L = 0$$

PediTools Website

The PediTools Web calculators [34] were developed in PHP, a general-purpose scripting language well suited to Web development [35]. The website was generated using RapidWeaver version 7.5.7 (Realmac Software) [36]. Web hosting is currently provided by Bluehost Inc [37], under a shared hosting environment. Access statistics are tracked via Google Analytics. The PediTools Web server is configured to require the use of Secure Sockets Layer to encrypt traffic to and from the Web server. No data entered as inputs for the medical calculators are saved or analyzed.

For the PediTools Electronic Growth Chart, a Microsoft Excel spreadsheet template was designed to allow users to enter protected health information locally, but it would calculate the nonprotected health information values needed to generate a growth chart. Specifically, the date of birth, gender, gestational age at birth, and specific dates and measurements are entered, but only the gender and calculated postmenstrual ages and anthropometric measures are submitted to the PediTools electronic growth chart site, avoiding transmission of any specific dates.

For the PediTools electronic growth chart, LMS-based calculations were performed as above for all the measurements submitted. In addition, for sequential measurements, rate of weight change in grams per week, both observed and expected (to maintain the previous growth centile), were displayed.

As a visual aid to recognize excessive loss of weight Z-score between sequential measurements, after the first 10 days of life, the change in weight Z-score (ΔZ) was color coded to display as red if the Z-score decreased by more than 0.06 SDs per week; yellow for decrease by more than 0.03 SD per week; and green otherwise. These thresholds were chosen somewhat arbitrarily, but over the course of a 14-week admission, each color would indicate an overall ΔZ weight of −0.84, −0.42, or less negative than −0.42 SDs, respectively.

Multisite Comparison of Growth Outcomes

The Vermont Oxford Network (VON) is a nonprofit voluntary collaboration of neonatal health care professionals representing more than 1200 hospitals around the world [38]. Deidentified data were obtained from 7 level 3 NICUs in Massachusetts, which participate in the VON registry. Eligibility criteria included birth year from 2006 to 2011; gestational age between 23 0/7 and 34 6/7 weeks; no severe congenital malformations; and survival to discharge. Availability of birth weight, discharge weight, and length of stay were required to calculate the weight Z-score at birth and discharge. Infants were excluded if birth or discharge weight Z-scores were less than −4 or greater than 4, as values beyond these extremes often reflected data entry error. The calculated outcome metric was the ΔZ weight from birth to discharge. For NICU C, data for neonates born between 2012 and 2017 were also obtained for postintervention quality improvement outcomes analysis.

The VON registry provides a manual of operations with data definitions and eNICQ software, which allows for the collection, error checking, and submission of infant data. These manuals, data collection forms, and electronic data submission instructions are all available on the VON website. At each hospital, individual patient-level data for that hospital were exported from eNICQ as a CSV file, with 1 row per patient and 1 column per data field. (As of 2019, eNICQ data exports are now in XML or JSON format, but they contain identical information.) The data columns abstracted for each hospital in this study included the following: birth year (BYEAR), initial gestational age (GAWEEKS, GADAYS), birth weight (BWGT), length of stay (LOS1), and discharge weight (DWGT). Additional information obtained included source of admission (inborn or outborn, LOCATE), day of life of admission (DAYADMISS), discharge disposition (home, transfer to another facility, or death, FDISP), and congenital malformations (CMAL). Gender was not obtained, as at the time the study was originally conceived, the only preterm growth chart with LMS parameters available (Fenton 2003) was not gender specific [19]. Outcomes obtained but not reported here included the following: birth (BHEADCIR) and discharge (DHEADCIR) head circumference, early (EBSEPS) or late (LBPATH) bacterial infection, oxygen requirement at 36 weeks postmenstrual age (OX36), necrotizing enterocolitis (NEC, NECSURG), and retinopathy of prematurity (ISTAGE, ROPSURG).

Analysis was performed using R, free software for statistical computing [39], using the free version of the RStudio integrated development environment [40]. For data visualizations, smoothed conditional mean curves were generated by the R ggplot [41] package via generalized additive model and cubic splines [42]. When present, the bands surrounding the smoothed curves represent the 95% CI around the mean.

Comparisons among hospitals were performed by 2-tailed *t* test, analysis of variance, Wilcoxon rank-sum, and Kruskal-Wallis test, as appropriate. Multihospital ΔZ analysis was performed by fitting a linear model of ΔZ versus gestational age, with interaction terms for both slope and intercept for each hospital. When multiple pairwise comparisons were performed, multiple testing adjustment was performed by the Tukey honestly significant difference method. The study was approved by the Institutional Review Boards at each of the hospitals that contributed data.

Single-Site Growth Outcomes Quality Improvement Project

Multiple bundled growth and nutrition quality improvement interventions were essentially simultaneously implemented at NICU C, starting in late 2011. These bundled changes included the following: (1) raised awareness of baseline growth failure by educational presentations to clinicians, showing how growth outcomes differed between NICU C and NICU F; (2) development of an electronic growth chart, as described in the PediTools Web tool; (3) systematic weekly growth metric collection in a form compatible with the electronic growth chart tool; (4) formal review of all NICU patients and their interval growth at weekly multidisciplinary rounds with pediatric dieticians; (5) earlier and broader initiation of parenteral nutrition with increased protein content and more rapid advancement; (6) revision of enteral feeding advancement protocols, including earlier initiation of gut priming (trophic feeds).

Assessment of the effect of the bundled interventions was performed similar to the analysis of growth outcomes described above: ΔZ weight from admission to discharge was calculated for each patient and the results were analyzed over different birth year epochs.

Dissemination of Methods for Large-Scale Analysis

The R code used for the calculation of anthropometric measure Z scores from LMS parameters was bundled [43] into the R peditools package, and this will be hosted on GitHub [44] and shared under the Massachusetts Institute of Technology (MIT) License. The peditools package can be installed using the devtools package [45], with the command install_github("jhchou/peditools").

All growth charts described in this work are supported by the R peditools package, with the exception of the Fenton 2013 growth chart [21], for which the LMS parameters are available from the author by license only. If the Fenton 2013 parameters become publicly available in the future, they will be added to the peditools R package. In the meantime, the Olsen 2010 [25] or gender nonspecific Fenton 2003 [19] charts can be used to analyze preterm growth.

Results

PediTools Website

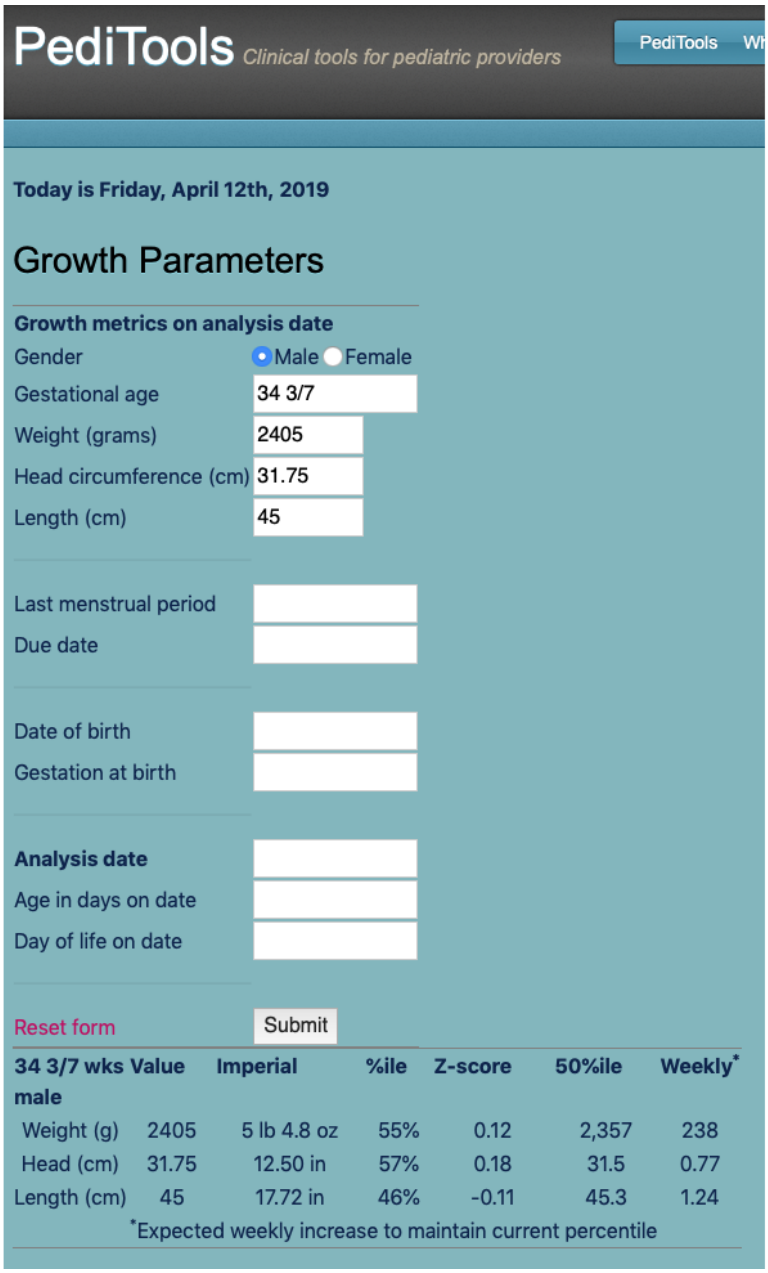
The first PediTools Web calculator was developed in 2011 as an in-house tool to improve documentation of anthropometric measurements of premature newborns cared for at hospital C by allowing calculation of Z-scores and percentiles, using the Fenton 2003 preterm growth chart [19], for which LMS parameters were published in 2007 [20]. The webpage was moved to public hosting in January of 2012. A screenshot of a representative Web-based PediTools growth calculator is shown in Figure 2.

Although accessible to the general public, the target audience and purpose of the PediTools website are pediatric clinical providers' bedside use. PediTools is agnostic to which growth charts are made available and does not provide recommendations as to which charts are appropriate for which populations. The users of the website are expected to exercise their own professional clinical judgment to determine suitability for their purposes.

Additional growth chart calculators have subsequently been added to PediTools, including support for the Fenton 2013 preterm chart [21], CDC infant and pediatric [22,23], World Health Organization (WHO) infant [24], Olsen 2010 preterm [25], WHO arm circumference and triceps and subscapular skinfold [26], CDC triceps and subscapular skinfold [27], Olsen 2015 BMI for preterm [28], Zemel 2015 Down syndrome [29,30], Abdel-Rahman 2017 midupper arm circumference [31], and Mramba 2017 midupper arm circumference [32].

As PediTools Web calculators were intended to be used by clinicians at the point of care, features in addition to reporting percentiles and Z-scores were integrated to promote ease of use and clinical relevance. For example, with the preterm calculators, a gestational age calculator was integrated to allow entry of either the postmenstrual gestational age of interest or any combination of last menstrual period, due date, delivery date, or chronologic age. For assessment of obesity, the CDC pediatric growth calculator includes both the Z score for BMI and updated categorization of extreme obesity, defined as BMI≥120% of the 95th percentile or ≥35 kg/m^2 [46]. Both international and imperial units are supported. For infant calculators, calculations at both chronologic age and age corrected for prematurity can be reported, which is beneficial when assessing for timely attainment of developmental milestones. To help set goals for future growth, the calculators report the expected amount that anthropometric measures should increase over time to maintain the current Z-score (ie, equivalent to growing along the current percentile curve).

Figure 2. Screenshot of a representative PediTools web-based growth calculator (Fenton 2013 for preterm infants). The upper section demonstrates flexible support for multiple methods of input data entry. Data entry can include age as either gestational age or specific dates; measurements can be entered in metric or imperial units (grams or pounds and ounces; centimeters or inches); and even if no measurement is entered, the expected median (50th percentile) values will be displayed. The lower section displays the results of the LMS-based calculations, including the anthropometric measures in both metric and imperial units, percentile, Z-score, expected median measurement for age, and weekly growth required to maintain the current percentile.



PediTools Website Utilization

The PediTools Web calculators have been well received by the clinical community. Since its public launch in January 2012, website access has increased to more than 500,000 page views per month. Figure 3 documents the increasing monthly page views over time; Table 2 shows page views by calculator for the year ending June 2019. Users were primarily from the United States (433,438/520,450; 83.28% users), but there were at least 1000 users from each of another 21 countries, with over 3000 users from Canada (17,169/520,450; 3.30%), India (5619/520,450; 1.08%), Australia (5096/520,450; 0.98%), Mexico (4066/520,450; 0.78%), and Brazil (3546/520,450; 0.68%). Access was primarily from desktop devices (307,326/518,796; 59.23%), followed by mobile devices (201,970/518,796; 38.93%) and tablets (9500/518,796; 1.83%).

PediTools also includes several aids not related to anthropometric measurements, including a bilirubin tool, which assists in the management of neonatal hyperbilirubinemia per the American Academy of Pediatrics 2004 guidelines [47,48] and a stand-alone version of the gestational age calculator, which is also incorporated in the preterm growth calculators. They will not be further discussed here, but they are mentioned as they receive the 4th and 6th largest number of page views, respectively.

Figure 3. PediTools (https://peditools.org/) website overall monthly pageviews over time from public launch in January 2012 through June 2019.

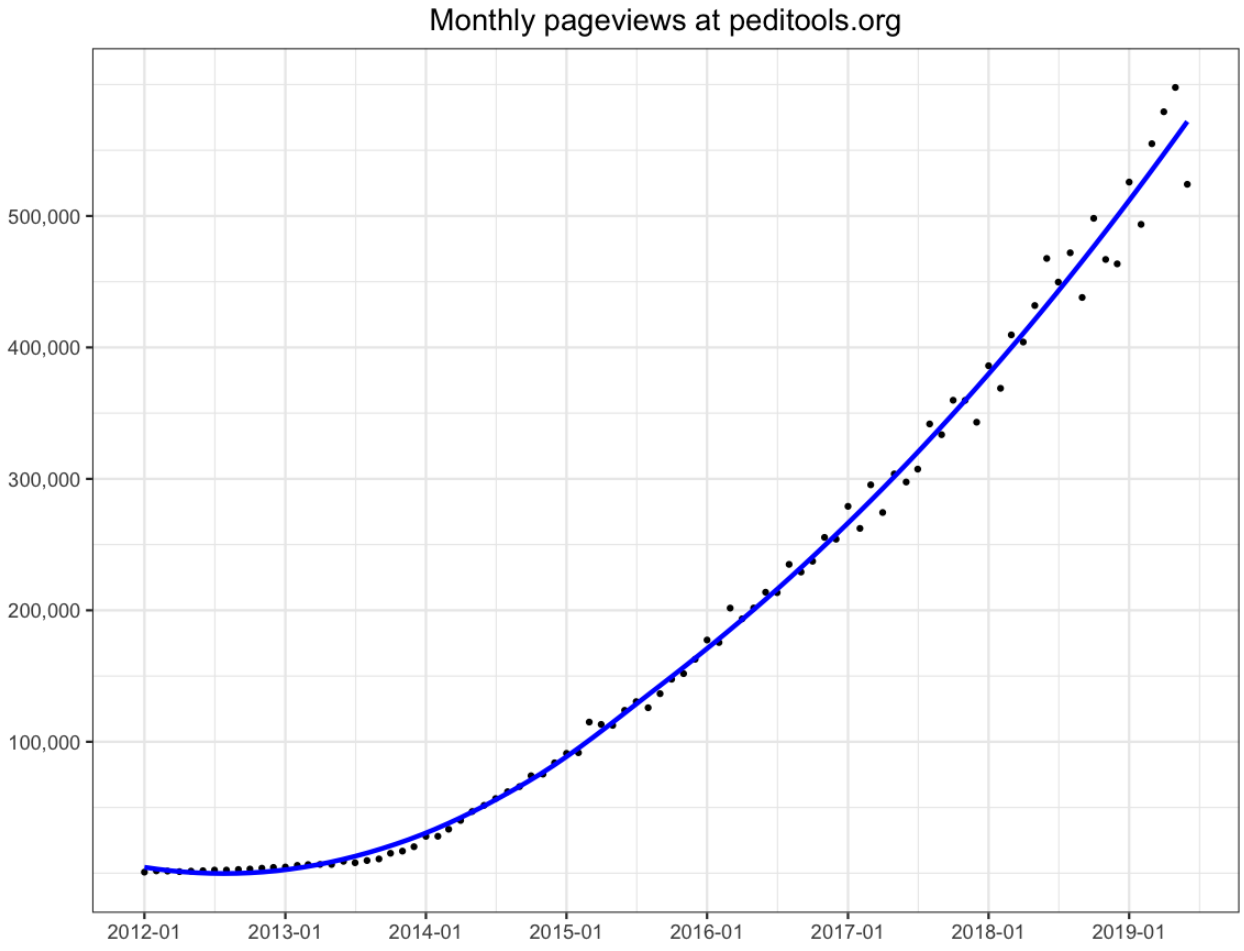


Table 2. PediTools page views by calculator for year ending June 2019.

Web page views	Value (N=5,192,170), n (%)
Fenton 2013 preterm	1,438,367 (27.54)
CDC[a] pediatric	1,338,920 (25.64)
WHO[b] infant	954,634 (18.28)
Bilirubin tool	411,897 (7.89)
CDC infant	360,440 (6.94)
Gestational age tool	257,201 (4.92)
Olsen 2010 preterm	126,486 (2.42)
CDC mid-upper arm circ	79,654 (1.53)
Electronic growth chart	79,051 (1.51)
Down syndrome, infant	46,163 (0.88)
Olsen BMI preterm	35,396 (0.68)
Down syndrome, pediatric	28,214 (0.54)
WHO arm and skinfold	21,185 (0.41)
Fenton 2003 preterm	14,562 (0.28)

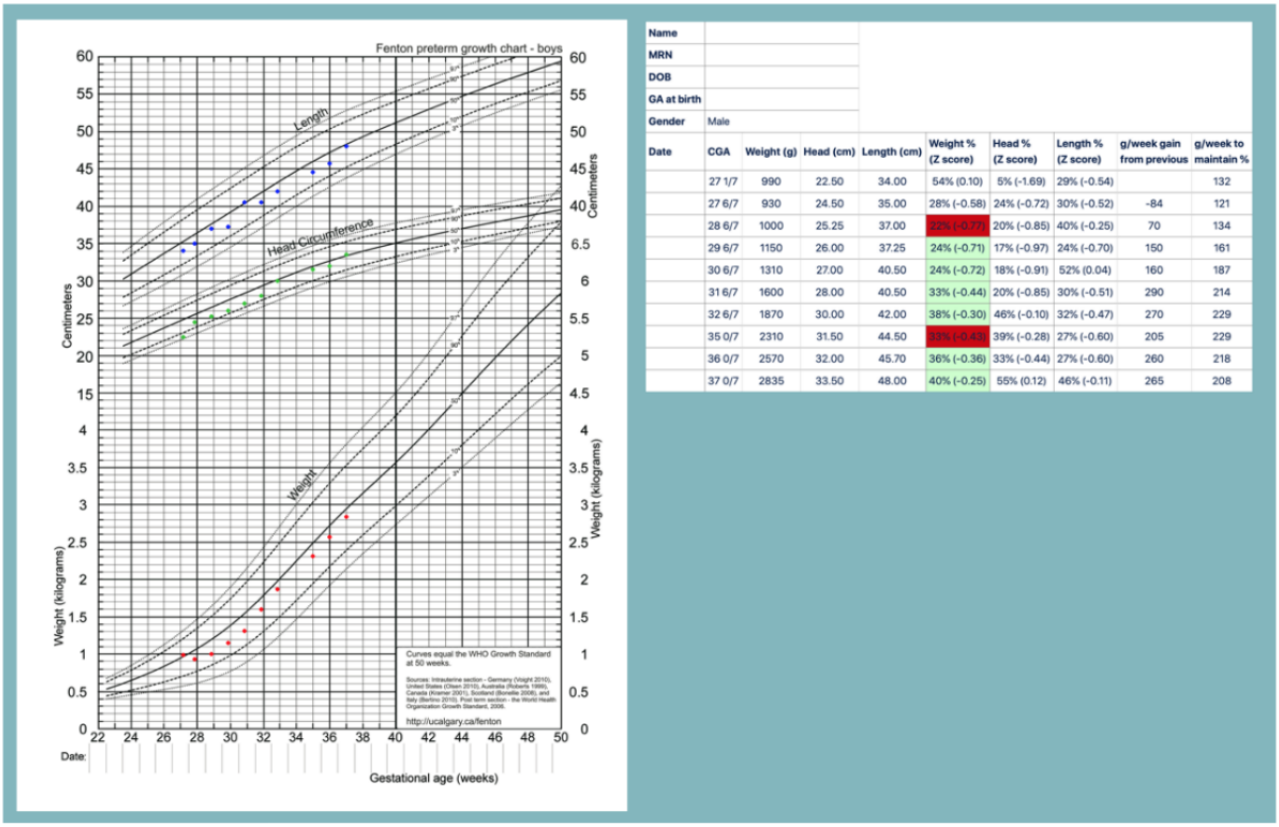
[a]CDC: Centers for Disease Control and Prevention.

[b]WHO: World Health Organization.

Electronic Growth Chart

The growth chart Web calculators on the PediTools website are limited in that only a single measurement can be analyzed at a time, whereas growth reflects how measurements change over time. For the Fenton 2013 preterm growth chart, an additional tool was developed to allow monitoring growth over time. As a public access website, care needs to be taken to not encourage sending protected health information over the internet. A Microsoft Excel spreadsheet was developed, in which specific dates and measures could be entered, but only deidentified data would be copied and pasted for secure submission via a webpage form. The output graphic was based on the original published chart [21] but with all the points plotted and supplemented with a table of percentiles, Z-scores, expected versus observed growth, and clinical decision support provided by color coding significant changes in Z-score between measurements (Figure 4). This tool was also used as part of a quality improvement project for longitudinal growth outcome monitoring (see below).

Figure 4. Electronic Fenton 2013 preterm growth chart. De-identified demographic and anthropometric data is copied into a webpage form from a specifically designed Microsoft Excel™ spreadsheet. The upper panel shows each anthropometric measurement plotted automatically onto the traditional paper-based chart. The lower panel displays calculated percentiles, Z-scores, and weekly weight change, both the actual observed change as well as the expected weekly change needed to maintain the previous percentile. Clinical decision support is provided by color-coding based on the weekly weight Z-score change.



Name									
MRN									
DOB									
GA at birth									
Gender	Male								
Date	CGA	Weight (g)	Head (cm)	Length (cm)	Weight % (Z score)	Head % (Z score)	Length % (Z score)	g/week gain from previous	g/week to maintain %
	27 1/7	990	22.50	34.00	54% (0.10)	5% (-1.69)	29% (-0.54)		132
	27 6/7	930	24.50	35.00	28% (-0.58)	24% (-0.72)	30% (-0.52)	-84	121
	28 6/7	1000	25.25	37.00	22% (-0.77)	20% (-0.85)	40% (-0.25)	70	134
	29 6/7	1150	26.00	37.25	24% (-0.71)	17% (-0.97)	24% (-0.70)	150	161
	30 6/7	1310	27.00	40.50	24% (-0.72)	18% (-0.91)	52% (0.04)	160	187
	31 6/7	1600	28.00	40.50	33% (-0.44)	20% (-0.85)	30% (-0.51)	290	214
	32 6/7	1870	30.00	42.00	38% (-0.30)	46% (-0.10)	32% (-0.47)	270	229
	35 0/7	2310	31.50	44.50	33% (-0.43)	39% (-0.28)	27% (-0.60)	205	229
	36 0/7	2570	32.00	45.70	36% (-0.36)	33% (-0.44)	27% (-0.60)	260	218
	37 0/7	2835	33.50	48.00	40% (-0.25)	55% (0.12)	46% (-0.11)	265	208

Multisite Comparison of Growth Outcomes

Variation in Overall Growth Outcomes at Different Hospitals

Our first aim was to demonstrate the feasibility of using ΔZ to assess growth outcomes of premature newborns and to compare outcomes among hospitals. All infants born between 23 0/7 and 34 6/7 weeks gestational age from 2006 to 2011 at 7 level 3 NICUs in Massachusetts, with VON registry data available and who survived to discharge, were analyzed for growth outcomes analysis (Table 3). Weight Z-scores at birth and discharge and the change in Z-score from birth to discharge were calculated for each individual patient.

As shown in Table 3, the mean ΔZ from birth to discharge differed significantly by site (P<.001), with the overall mean ΔZ across all sites −0.81 and ranging across the 7 sites from −0.56 and −1.05.

Table 3. Study population of 7975 premature newborns born between 23 and 34 weeks of completed weeks gestation in 7 Massachusetts newborn intensive care units (A-G).

Metric	All NICUs[a] combined (n=7975)	A (n=461)	B (n=1586)	C[b] (n=1068)	D (n=418)	E (n=598)	F (n=1081)	G (n=2763)	*P* value
Gestational age (weeks), median (IQR)	32 (29.29-33.86)	29 (27.14-30.57)	33 (30.43-34.14)	32.86 (30.71-34.04)	31.86 (29-33.57)	29.14 (26.89-30.71)	31 (28.57-33.29)	32.57 (30.29-34)	<.001
Birth weight (gram), median (IQR)	1580 (1180-2050)	1130 (885-1316)	1810 (1370-2160)	1818.50 (1380-2195)	1462.50 (1100-1943)	1160 (871.25-1365)	1400 (1065-1860)	1750 (1340-2100)	<.001
Birth weight Z-score, mean (SD)	−0.19 (0.84)	−0.40 (0.85)	−0.14 (0.83)	−0.07 (0.81)	−0.32 (0.79)	−0.36 (0.91)	−0.24 (0.85)	−0.14 (0.82)	<.001
Discharge postmenstrual age (weeks), mean (SD)	36.4 (2.86)	35.71 (4.37)	36.64 (2.22)	36.17 (2.42)	34.88 (3.03)	36.95 (2.70)	37.05 (3.60)	36.31 (2.58)	<.001
Discharge weight Z-score, mean (SD)	−1.00 (0.80)	−1.15 (0.81)	−1.19 (0.77)	−0.88 (0.78)	−0.90 (0.76)	−1.32 (0.77)	−0.80 (0.80)	−0.93 (0.79)	<.001
Weight delta Z[c], mean (SD)	−0.81 (0.52)	−0.75 (0.49)	−1.05 (0.49)	−0.80 (0.40)	−0.58 (0.46)	−0.96 (0.59)	−0.56 (0.52)	−0.79 (0.49)	<.001

[a]NICU: neonatal intensive care unit.

[b]An additional 1120 neonates born between 2012 and 2017 from NICU C were included for postintervention outcomes analysis, not tabulated here.

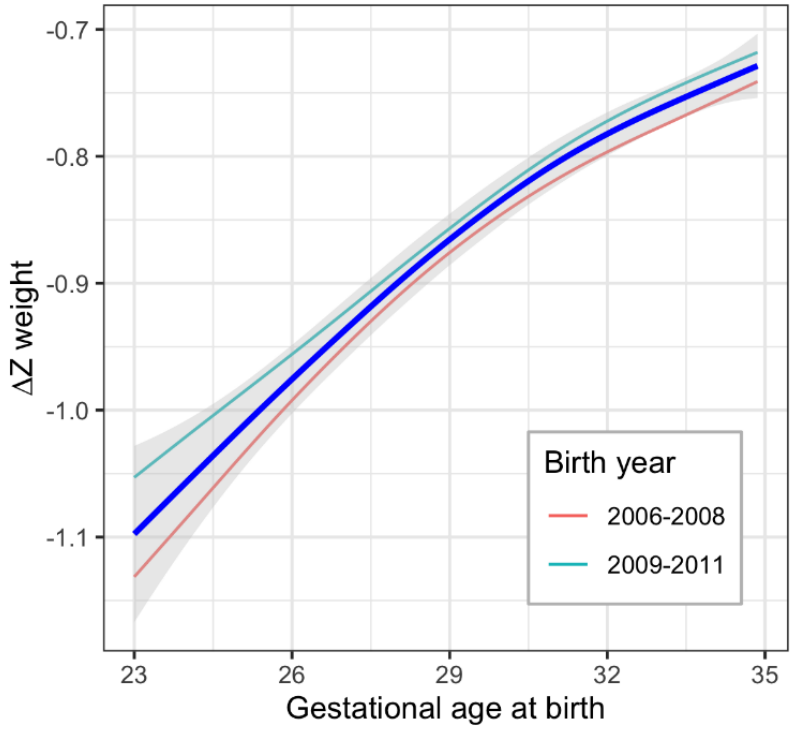
[c]Weight delta Z is the change in Z-score for weight from birth to discharge [19,20].

Correlation of Growth Failure With Gestational Age at Birth

Growth outcomes in the preterm population are potentially dependent on multiple factors, including gestational age at birth, nutrition practices, and timing of discharge. One of our hypotheses was that the ΔZ weight from birth to discharge might be associated with gestational age at birth. Combining data from all 7 hospitals across the entire time period from 2006 to 2011 and plotting the ΔZ weight versus gestational age at birth showed inferior growth (ie, more negative change in Z-score) with increasing prematurity (Figure 5). Grouping the data by birth year 2006 to 2008 and 2009 to 2011 showed that the relationship between more negative ΔZ weight and lower gestational ages appears unchanged over the 2 time epochs, although the shift in the lines upward suggests less loss in weight ΔZ score in the later epoch.

By visual inspection, the relationship between ΔZ weight and gestation at birth appears roughly linear. Fitting a linear regression allowed estimation of the relationship between growth failure and gestation at birth. At 29 0/7 weeks, the expected mean ΔZ weight from birth to discharge was −0.88 (P<.001, 95% CI −0.865 to −0.893), with each additional week of decrease in gestational age contributing an additional −0.029 (P<.001, 95% CI −0.025 to −0.033).

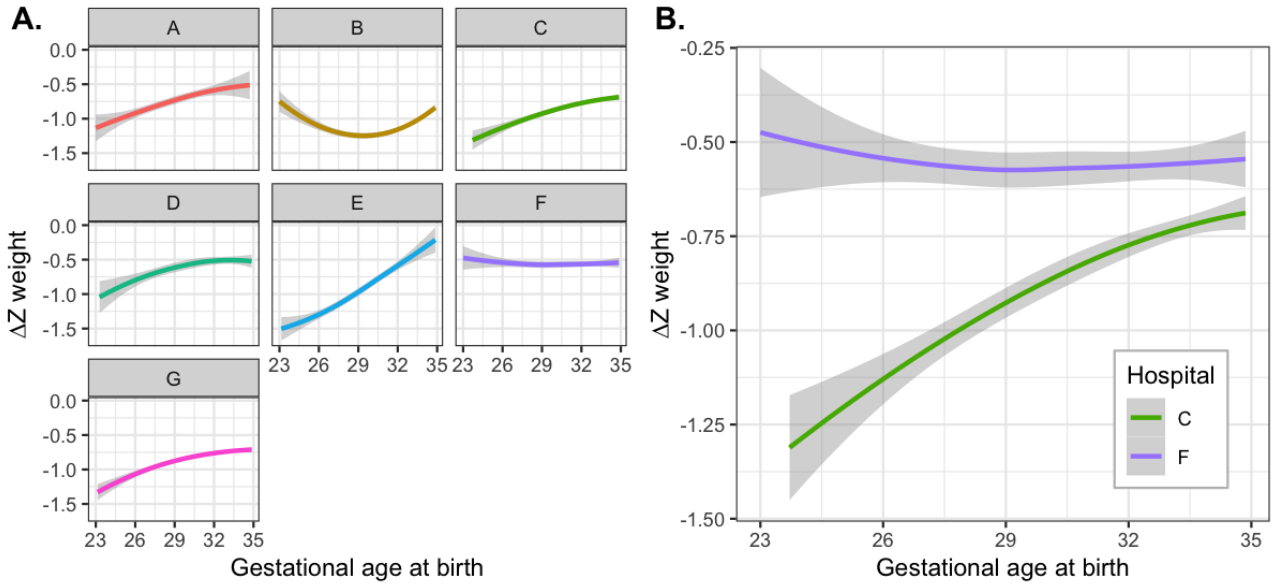
Figure 5. Change in weight Z-score from birth to discharge versus gestational age at birth, demonstrating inferior growth with increasing prematurity for all seven NICUs combined. The dark blue line is for all years 2006 - 2011 combined with the gray band representing the 95% confidence interval; the thin lines show the grouped birth years 2006 - 2008 versus 2009 - 2011.



Interhospital Variation Between Growth Outcomes and Gestational Age at Birth

It was possible that this inverse relationship between growth failure and gestational age at birth was intrinsic to prematurity and was therefore universal among the hospitals. To test this hypothesis, we next analyzed whether different hospitals might have different growth outcome characteristics. We found that the relationship between ΔZ weight and gestation at birth differed by hospital (Figure 6), with significant interhospital variation in both the degree of growth failure and the interaction with gestational age at birth. Some hospitals show much inferior growth at lower gestational ages at birth (eg, hospital E), whereas other hospitals show better growth overall and absence of inferior growth at lower gestational age (eg, hospital F). Patterns of growth at individual hospitals remained stable across different birth year epochs (data not shown), suggesting reliability for use as a quality improvement metric.

Figure 6. Inter-hospital variation in change in weight Z-score from birth to discharge, as related to gestational age at birth, (A) separately for each of seven different hospital NICUs in Massachusetts, and (B) for hospitals C and F overlaid on the same plot to better demonstrate inter-hospital differences.



The differences can be seen clearly when choosing hospital F as the baseline hospital and comparing pairwise with every other hospital (eg, Figure 6). Fitting a linear model of ΔZ versus gestational age, with interaction terms for both slope and intercept for each hospital, demonstrated statistically significant differences between hospitals A-E and G compared with hospital F as baseline. Hospital F did not show a relationship between gestational age at birth and ΔZ weight (slope=−0.001, *P*=.85, 95% CI −0.010 to 0.009), and at 29 0/7 weeks, the mean ΔZ was −0.553 (*P*<.001, 95% CI −0.586 to −0.520). In comparison with hospital F, each of the other hospitals had both a more negative ΔZ at 29 0/7 weeks (all *P*<.003) and a greater relationship between increasing prematurity and more negative ΔZ (all *P*<.001).

In summary, we found that although all 7 hospitals studied had negative weight ΔZ from birth to discharge, hospitals differed in degree of negative ΔZ (ie, intercept at a gestational age of 29 0/7 weeks), as well as the degree to which inferior growth was related to lower gestational ages (ie, slope).

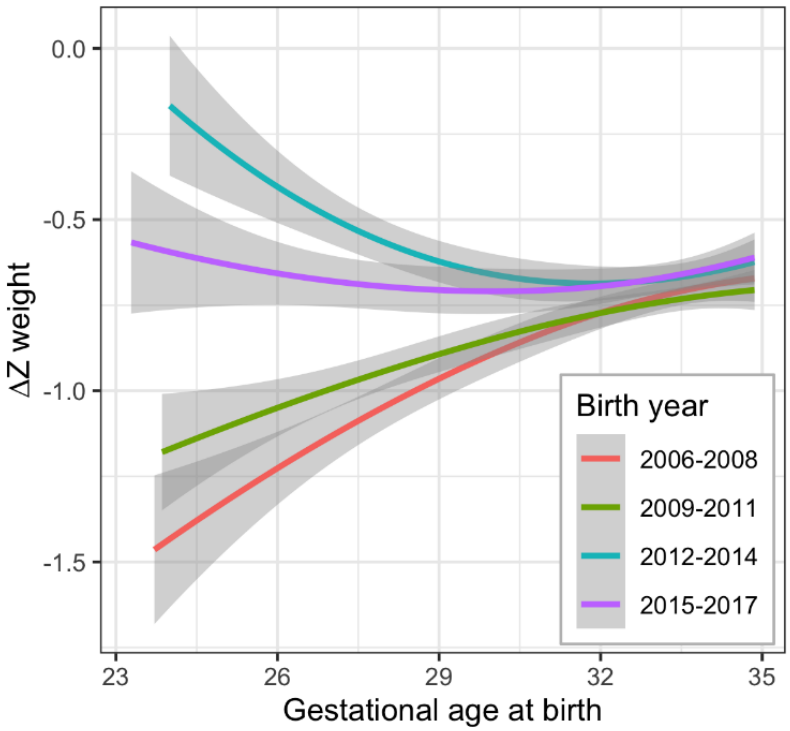
Single-Site Growth Outcomes Quality Improvement Project

As poor growth trajectory might exacerbate long-term neurodevelopmental impairment, particularly for those patients born the most preterm, NICU C embarked on a multifocal quality improvement project to reduce the loss in weight Z-score from birth to discharge. Bundled interventions introduced in 2011 targeting factors potentially contributing to poor growth included the following:

- Utilized baseline data to raise awareness of poor growth, for example, in comparison with NICU F
- Implemented system of weekly growth metric collection
- Formal weekly multidisciplinary (including pediatric dietician) review of electronic growth chart
- Earlier and broader initiation of parenteral nutrition
- Increased protein content in premade parenteral nutrition
- Accelerated advancement of parenteral nutrition
- Earlier initiation of enteral nutrition
- Revised enteral feeding advancement protocol
- Pasteurized donor human milk made available

These interventions were associated with a significant reduction in loss of weight Z-score, particularly at the lowest gestational ages, which was sustained and progressive (Figure 7). Compared with the preintervention epoch 2006 to 2008, there was a marked reduction in loss of birth weight Z-score in the postintervention epoch from 2012 to 2014.

Figure 7. Improvement of growth outcomes (ΔZ weight) at hospital C, by birth gestation and birth year epoch. Epoch 2006 – 2008 is pre-intervention; 2009 - 2011 covers the beginning of implementation of interventions; 2012 - 2014 is the immediate post-intervention epoch; 2015 - 2017 demonstrates sustained improvement, but less extreme at lower gestational ages, after targeting a goal ΔZ weight of -0.6. The largest improvements are seen at the lowest gestational ages at birth.



Using the 2006 to 2008 epoch as a reference and fitting a linear model of ΔZ weight versus gestational age at birth, with interaction terms for both slope and intercept (centered at 29 weeks) for the remaining 3 epochs of 2009 to 2011, 2012 to 2014, and 2015 to 2017, showed significant differences in both the ΔZ weight at 29 weeks (differences in intercept, *P*=.03, <.001, and <.001) and association with gestational age at birth (differences in slope, *P*=.015, <.001, and <.001, respectively).

In fact, because of concern that the degree of reduction of weight Z-score loss at the lowest gestational ages might be excessive in the 2012 to 2014 epoch, potentially contributing to future development of the metabolic syndrome [12], a less aggressive

approach was adopted. The subsequent growth target was a ΔZ from birth to discharge of roughly −0.6 for all gestations, resulting in the more flattened curve of epoch 2015 to 2017 (Figure 7).

Dissemination of Methods for Large-Scale Analysis

The PediTools website has met a need for clinicians wishing to analyze data for an individual patient at a time. In contrast, the multisite comparison of growth outcomes of thousands of patients at 7 NICUs yielded additional insight into growth patterns of preterm newborns and prompted a successful quality improvement project at 1 NICU. Here, we describe the release of open-source computer code to permit others to conveniently do similar analysis, which should be useful for both clinical research and quality improvement monitoring.

The peditools R package will be made available on GitHub and provide functions to work with LMS-based anthropometric charts, including all LMS parameters possible [44]. All growth charts available on the PediTools website (listed in Table 1), with the exception of the Fenton 2013 preterm growth chart, are included. At this time, the Fenton 2013 chart LMS parameters are available by license only and are restricted from being shared. As additional charts are added to the PediTools website, the peditools R package will also be updated to include the new charts. The peditools R package will be released under the relatively permissive MIT License, which allows for commercial use, modification, redistribution, and sublicensing.

The primary package tool is the peditools::x_to_z() function, which takes as inputs a vector of anthropometric measurements, a vector of ages, a vector of genders, and a uniquely specified chart and measure, and which outputs a vector of Z-scores. In addition, helper functions peditools::recode_von() and peditools::recode_von_xml() are included to easily import VON datasets (exported as a CSV or XML file) into the R environment for analysis.

Discussion

In this work, we discuss the benefits of developing software tools to perform calculations on LMS-based growth charts and present examples of their utilization in patient care delivery, clinical research, and quality improvement projects.

Principal Findings and Limitations

PediTools Website

The publicly accessible PediTools website [34] makes possible the calculation of exact Z-scores and percentiles for 13 distinct growth charts. Despite the availability of published paper forms of these charts and many of the LMS parameters, there appears to have been an unmet need for publicly available calculators, as demonstrated by PediTools page views increasing to over 500,000 per month. Most visitors (433,438/520,450; 83.28%) are from the United States, but 21 countries had at least 3000 distinct users in the previous year. It is likely that most visitors are health care providers, as inspection of the 100 service provider networks with the largest number of PediTools access sessions in the past year revealed that 52 of the network names contained one of the words health, health care, hospital, or medical. In addition, most email communications to PediTools support have been from dieticians, with some from physicians.

Other than documenting website access statistics, it is difficult to gauge the degree of clinical and research impact of the PediTools website, as before this publication, no citable reference or digital object identifier has been available to allow citation tracking. However, in a nonexhaustive internet search, the PediTools website itself is cited in a number of publications, reviews, and clinical guidelines related to topics such as identifying neonatal and pediatric malnutrition, neurodevelopmental outcomes of preterm newborns, bariatric surgery guidelines, nutrition delivery in chronic disease, and monitoring of postnatal growth in late-preterm newborns [2,49-56]. The combination of website access statistics and citations suggests that the suite of PediTools calculators provides a useful service to practicing clinicians.

PediTools is primarily accessed by users in the United States. It is unclear whether clinicians in other countries use other tools, perhaps localized to their specific populations [57,58]. Alternatively, there may be lack of awareness of the tools' availability. Dissemination of PediTools has thus far been entirely by word of mouth, and its development has thus far neither been formally presented at conferences nor previously published.

A limitation is that the calculations performed by PediTools are all done server side; therefore, in areas with limited internet availability, the tools are inaccessible. Work is in progress to develop a number of the tools as mobile device apps that do not require internet connectivity, with some preliminary work on iOS now released [59,60].

Another limitation of the PediTools website is that, currently, only charts with LMS-based parameterizations are offered. In some instances, LMS parameterization has been done, but the parameters are not published [61]. Alternative methods of parameterization have also been utilized [62], for example, quantile regression for nomogram generation [63] or fitting a skew t-distribution [64]. The PediTools calculators were implemented in PHP, which works well as a general-purpose scripting language, but it does not generally support more complex statistical calculations. For example, the skew t-distribution does not have a closed form solution, but specialized software in other languages (eg, the GAMLSS package in R) [65] would allow calculation of exact Z-scores, given the model's 4 parameters (mu, sigma, nu, and tau). A future extension of the PediTools R package could incorporate calculations for charts utilizing different parameterizations.

Multisite Comparison of Growth Outcomes in Preterm Infants

The PediTools website analyzes a single patient at a time, as might be appropriate for management of individual patients. Upon applying LMS-based calculations to a large cohort of infants from 7 hospitals in Massachusetts, we were able to characterize the ΔZ weight. Across the overall population, findings included a significant decrease in weight Z-score and an association with larger decreases in Z-score at lower gestational ages. When each of the 7 hospitals was analyzed

separately and compared, we found significant interhospital variation in decreases in Z-score and in the degree of association with gestational age. The findings remained similar across different birth year epochs (Figure 5 and data not shown). This observation of growth outcomes is potentially concerning, as growth failure in this vulnerable population is associated with poor neurodevelopmental outcomes, and we show here that the infants at highest risk of poor neurodevelopmental outcome—those born the most preterm—are also at greatest risk of poor growth.

For a number of reasons, the interhospital variations should be taken as a proof of concept and feasibility demonstration of the approach rather than a rigorous comparative analysis of the 7 hospitals. The patient populations of the hospitals differed significantly (Table 3). Although all 7 hospitals participated in the VON registry, which served as the source of the data, participation varied among hospitals, with the Very Low Birth Weight database (401-1500 grams birth weight or 22-29 weeks completed gestation), the Expanded database (all infants admitted to a NICU within 28 days of birth), or even changing participation during the time period of this study. In addition, the discharge disposition varied from 6.4% to 70% transfer to another hospital, versus discharge home. No effort was made to document differences in nutritional practices at each institution. That being said, reanalysis of the dataset restricted to either requiring birth weight<1500 grams or requiring a home discharge disposition did not substantially change any of the findings reported here (data not shown), suggesting that the differences observed were robust to these varied patient populations.

Single-Site Growth Outcomes Quality Improvement Project

Assessing outcomes at the hospital level may help identify specific practice differences effective in improving growth, as well as providing a metric to assess and follow performance. In this report, a hospital implemented a bundle of interventions and utilized LMS-based assessment of ΔZ weight from birth to discharge by gestational age at birth to monitor the impact pre- and postintervention. Not only was this method helpful in showing statistically significant changes in improvement in overall growth and reducing the impact of lower gestational age on inferior growth but it was also useful in helping to recognize possible excessive growth (eg, in the most preterm infants in the 2012-2014 epoch).

As a quality improvement project, there was less emphasis on attempting to delineate which specific changes in practice had the greatest impact on outcome, and there was more emphasis on rigorous monitoring of the effect of implementing multiple potentially better practices. We believe that the greatest impact likely came from the consistent, weekly, multidisciplinary review of the ongoing growth of each and every patient in the NICU, as well as ongoing monitoring of neonatal growth as a unit-wide metric. The use of LMS-based calculation of exact Z-scores was critical for this intervention.

A challenge in targeting growth in preterm infants is the lack of evidence conclusively demonstrating exactly what ideal growth should be, but consensus guidelines are emerging [6,49]. Identified indicators of malnutrition include the following: ΔZ over time (with goal ΔZ weight not more negative than −0.8, roughly matching hospital C's goal after 2014 of ΔZ weight=−0.6), weight gain velocity, actual nutrient intake, days to regain birth weight, length growth velocity, and ΔZ of length for age. A major purpose of the PediTools LMS-based calculators was to make these data easy for clinicians to analyze, track, and understand.

Comparison With Previous Work

Previous work has analyzed growth of large populations of preterm newborns. Horbar et al [66] drew on data obtained from the full Vermont Oxford Registry on 362,833 newborns born between 2000 and 2013, with birth weight from 501 to 1500 grams. In this large, aggregate population, they reported improvements of growth velocity and a decrease in discharge with growth failure and severe growth failure (defined as discharge at less than the 10th and 3rd percentiles), across the time period from 2000 to 2013. Similarly, Griffin et al [67] reported on 25,899 infants born in California, with birth weight from 500 to 1500 grams or gestational age from 22 to 32 weeks, born between 2005 and 2012. They demonstrated a reduction in fall in weight Z-score between birth and discharge over the time period, as well as a reduction in the proportion of infants discharged home below the 10th percentile for weight or ΔZ weight less than −1. We see similar improvements in less negative ΔZ weight over time, comparing birth years 2006 to 2008 versus 2009 to 2011 (Figure 5).

Although both Horbar et al [66] and Griffin et al [67] report outcomes by birth weight (binned into categorical groups of 250 gram increments and which would therefore include both large for gestation more premature and small for gestation less premature newborns), previous studies have not reported the association described here between growth outcomes as ΔZ weight and gestational age at birth (both continuous variables).

Both of these studies report on the important findings of overall population-level improvement in growth outcomes in preterm newborns across a time period ranging from 2000 to 2013, which likely reflects clinical practice changes across the field of neonatology as a whole, which may result in improved long-term outcomes. However, this information is less helpful for clinicians attempting to assess outcomes at the local hospital level. In addition, the previous studies have not shared the tools needed to make it convenient to perform this analysis on new populations. In fact, without tools to easily assess growth, it is not easy for clinicians to even recognize that there might be an issue with growth outcomes in their patient populations.

A major goal of this study is to make tools available, allowing others to perform their own large-scale growth outcomes analysis, facilitating future research to better describe ideal growth that will lead to optimal long-term outcomes. More than 1200 hospitals around the world participate in the VON registry, and an increasing number of clinical sites use electronic health records from which anthropometric data can be extracted. With this increased availability of growth data, clinicians can easily replicate this analysis in a very small amount of code, using free and open-source tools, including the R statistical programming language, the RStudio integrated development

environment, the ggplot2 R visualization package [39-41], and the peditools package described here [44].

Conclusions

Tools to perform LMS-based growth chart calculations have been made available on a public website and are highly utilized by clinical caregivers worldwide. Applying these methods to a large population of preterm newborns demonstrated widespread overall loss in weight Z-score from birth to discharge; that the magnitude of loss was associated with increasing prematurity, the population at the highest risk of poor neurodevelopment outcomes; and that there was significant interhospital variation in growth outcomes. At 1 site, these tools provided a convenient and reliable outcome measure for a clinical quality improvement project targeting growth. With this report, release of open-source code that implements LMS-based calculations will allow other clinicians and investigators to conveniently perform similar analyses with the promise to improve long-term outcomes in these high-risk pediatric patients.

Acknowledgments

This study was not externally funded. The authors wish to thank Drs Terri Gorman, Munish Gupta, Nneka Nzegwu, Sunita Pereira, and Alan Picarillo for contributing patient data. The authors also thank the many NICU caregivers for their commitment to excellent care of newborns and, particularly, dieticians Meghan Klauer and Charna Simon.

Conflicts of Interest

JHC is the owner of PediTools, LLC. As of February 2018, the PediTools website generates revenue from advertisements served by Google Adsense; Google Adsense has no input on the content presented on PediTools.

References

1. American Academy of Pediatrics, Barness LA. Pediatric Nutrition Handbook. Sixth Edition. Elk Grove Village, IL: American Academy of Pediatrics; 2009.
2. Green Corkins K, Teague EE. Pediatric nutrition assessment: anthropometrics to zinc. Nutr Clin Pract 2017 Feb;32(1):40-51. [doi: 10.1177/0884533616679639] [Medline: 30865347]
3. Kerac M, Frison S, Connell N, Page B, McGrath M. Informing the management of acute malnutrition in infants aged under 6 months (MAMI): risk factor analysis using nationally-representative demographic & health survey secondary data. PeerJ 2019;6:e5848 [FREE Full text] [doi: 10.7717/peerj.5848] [Medline: 31024756]
4. Younge NE, Newgard CB, Cotten CM, Goldberg RN, Muehlbauer MJ, Bain JR, et al. Disrupted maturation of the microbiota and metabolome among extremely preterm infants with postnatal growth failure. Sci Rep 2019 Jun 3;9(1):8167 [FREE Full text] [doi: 10.1038/s41598-019-44547-y] [Medline: 31160673]
5. Picon MM, Ruiz DM, Colciencias JI, Salcedo F, Colciencias JI, Alvis N. Risk factors of chronic infant malnutrition in Cartagena, Colombia: a case-control study. Pediatrics 2018 Jan;141:478 [FREE Full text] [doi: 10.1542/peds.141.1_MeetingAbstract.478]
6. Villar J, Giuliani F, Barros F, Roggero P, Zarco IA, Rego MA, et al. Monitoring the postnatal growth of preterm infants: a paradigm change. Pediatrics 2018 Feb;141(2):pii: e20172467 [FREE Full text] [doi: 10.1542/peds.2017-2467] [Medline: 29301912]
7. Andrews ET, Ashton JJ, Pearson F, Beattie RM, Johnson MJ. Early postnatal growth failure in preterm infants is not inevitable. Arch Dis Child Fetal Neonatal Ed 2019 May;104(3):F235-F241. [doi: 10.1136/archdischild-2018-315082] [Medline: 30135111]
8. Scherdel P, Dunkel L, van Dommelen P, Goulet O, Salaün JF, Brauner R, et al. Growth monitoring as an early detection tool: a systematic review. Lancet Diabetes Endocrinol 2016 May;4(5):447-456. [doi: 10.1016/S2213-8587(15)00392-7] [Medline: 26777129]
9. Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990 Jan;44(1):45-60. [Medline: 2354692]
10. Ramel SE, Demerath EW, Gray HL, Younge N, Boys C, Georgieff MK. The relationship of poor linear growth velocity with neonatal illness and two-year neurodevelopment in preterm infants. Neonatology 2012;102(1):19-24. [doi: 10.1159/000336127] [Medline: 22441508]
11. Ramel SE, Georgieff MK. Preterm nutrition and the brain. World Rev Nutr Diet 2014;110:190-200. [doi: 10.1159/000358467] [Medline: 24751630]
12. Ong KK, Kennedy K, Castañeda-Gutiérrez E, Forsyth S, Godfrey KM, Koletzko B, et al. Postnatal growth in preterm infants and later health outcomes: a systematic review. Acta Paediatr 2015 Oct;104(10):974-986 [FREE Full text] [doi: 10.1111/apa.13128] [Medline: 26179961]
13. Castanys-Muñoz E, Kennedy K, Castañeda-Gutiérrez E, Forsyth S, Godfrey KM, Koletzko B, et al. Systematic review indicates postnatal growth in term infants born small-for-gestational-age being associated with later neurocognitive and metabolic outcomes. Acta Paediatr 2017 Aug;106(8):1230-1238 [FREE Full text] [doi: 10.1111/apa.13868] [Medline: 28382722]

14. Ehrenkranz RA, Dusick AM, Vohr BR, Wright LL, Wrage LA, Poole WK. Growth in the neonatal intensive care unit influences neurodevelopmental and growth outcomes of extremely low birth weight infants. Pediatrics 2006 Apr;117(4):1253-1261. [doi: 10.1542/peds.2005-1368] [Medline: 16585322]
15. Shah PS, Wong KY, Merko S, Bishara R, Dunn M, Asztalos E, et al. Postnatal growth failure in preterm infants: ascertainment and relation to long-term outcome. J Perinat Med 2006;34(6):484-489. [doi: 10.1515/JPM.2006.094] [Medline: 17140299]
16. Franz AR, Pohlandt F, Bode H, Mihatsch WA, Sander S, Kron M, et al. Intrauterine, early neonatal, and postdischarge growth and neurodevelopmental outcome at 5.4 years in extremely preterm infants after intensive neonatal nutritional support. Pediatrics 2009 Jan;123(1):e101-e109. [doi: 10.1542/peds.2008-1352] [Medline: 19117831]
17. Sammallahti S, Pyhälä R, Lahti M, Lahti J, Pesonen A, Heinonen K, et al. Infant growth after preterm birth and neurocognitive abilities in young adulthood. J Pediatr 2014 Dec;165(6):1109-15.e3. [doi: 10.1016/j.jpeds.2014.08.028] [Medline: 25262301]
18. Zozaya C, Díaz C, de Pipaón MS. How should we define postnatal growth restriction in preterm infants? Neonatology 2018;114(2):177-180. [doi: 10.1159/000489388] [Medline: 29920494]
19. Fenton TR. A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format. BMC Pediatr 2003 Dec 16;3:13 [FREE Full text] [doi: 10.1186/1471-2431-3-13] [Medline: 14678563]
20. Fenton TR, Sauve RS. Using the LMS method to calculate z-scores for the Fenton preterm infant growth chart. Eur J Clin Nutr 2007 Dec;61(12):1380-1385. [doi: 10.1038/sj.ejcn.1602667] [Medline: 17299469]
21. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. BMC Pediatr 2013 Apr 20;13:59 [FREE Full text] [doi: 10.1186/1471-2431-13-59] [Medline: 23601190]
22. Centers for Disease Control and Prevention (CDC). 2017. Growth Charts - Percentile Data Files with LMS Values URL: https://www.cdc.gov/growthcharts/percentile_data_files.htm [accessed 2019-11-28]
23. Flegal KM, Cole TJ. Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. Natl Health Stat Report 2013 Feb 11(63):1-3 [FREE Full text] [Medline: 24992748]
24. Centers for Disease Control and Prevention (CDC). Growth Charts - WHO Child Growth Standards URL: https://www.cdc.gov/growthcharts/who_charts.htm [accessed 2019-11-28]
25. Olsen IE, Groveman SA, Lawson ML, Clark RH, Zemel BS. New intrauterine growth curves based on United States data. Pediatrics 2010 Feb;125(2):e214-e224. [doi: 10.1542/peds.2009-0913] [Medline: 20100760]
26. World Health Organization. WHO Child Growth Standards: Head Circumference-for-age, Arm Circumference-for-age, Triceps Skinfold-for-age and Subscapular Skinfold-for-age: Methods and Development. Geneva: World Health Organization; 2007.
27. Addo OY, Himes JH. Reference curves for triceps and subscapular skinfold thicknesses in US children and adolescents. Am J Clin Nutr 2010 Mar;91(3):635-642. [doi: 10.3945/ajcn.2009.28385] [Medline: 20053877]
28. Olsen IE, Lawson ML, Ferguson AN, Cantrell R, Grabich SC, Zemel BS, et al. BMI curves for preterm infants. Pediatrics 2015 Mar;135(3):e572-e581 [FREE Full text] [doi: 10.1542/peds.2014-2777] [Medline: 25687149]
29. Zemel BS, Pipan M, Stallings VA, Hall W, Schadt K, Freedman DS, et al. Growth charts for children with down syndrome in the United States. Pediatrics 2015 Nov;136(5):e1204-e1211 [FREE Full text] [doi: 10.1542/peds.2015-1652] [Medline: 26504127]
30. Centers for Disease Control and Prevention. 2017. Growth Charts for Children with Down Syndrome URL: https://www.cdc.gov/ncbddd/birthdefects/downsyndrome/growth-charts.html [accessed 2019-11-28]
31. Abdel-Rahman SM, Bi C, Thaete K. Construction of lambda, mu, sigma values for determining mid-upper arm circumference z scores in US children aged 2 months through 18 years. Nutr Clin Pract 2017 Feb;32(1):68-76. [doi: 10.1177/0884533616676597] [Medline: 27856693]
32. Mramba L, Ngari M, Mwangome M, Muchai L, Bauni E, Walker AS, et al. A growth reference for mid upper arm circumference for age among school age children and adolescents, and validation for mortality: growth curve construction and longitudinal cohort study. Br Med J 2017 Aug 3;358:j3423 [FREE Full text] [doi: 10.1136/bmj.j3423] [Medline: 28774873]
33. Abramowitz M, Stegun IA. Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. New York, NY: Dover; 2013.
34. PediTools. URL: https://peditools.org/ [accessed 2019-11-29]
35. PHP: Hypertext Preprocessor. URL: https://www.php.net/ [accessed 2019-11-28]
36. Realmac Software. RapidWeaver URL: https://www.realmacsoftware.com/rapidweaver/ [accessed 2019-11-28]
37. Bluehost. URL: https://www.bluehost.com/ [accessed 2019-11-28]
38. Vermont Oxford Network Registry. URL: https://public.vtoxford.org/ [accessed 2019-11-28]
39. The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. URL: https://www.R-project.org/ [accessed 2019-11-28]
40. RStudio. Boston, MA: RStudio, Inc; 2016. URL: http://www.rstudio.com/ [accessed 2019-11-28]
41. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
42. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. New York, NY: Springer; 2009.
43. Wickham H. R Packages. First Edition. Sebastopol, CA: O'Reilly Media; 2015.

44. GitHub. URL: https://github.com/jhchou/peditools [accessed 2019-11-29]
45. Wickham H, Hester J, Chang W. Devtools: Tools to Make Developing R Packages Easier. URL: https://devtools.r-lib.org [accessed 2019-11-28]
46. Styne DM, Arslanian SA, Connor EL, Farooqi IS, Murad MH, Silverstein JH, et al. Pediatric obesity-assessment, treatment, and prevention: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2017 Mar 1;102(3):709-757 [FREE Full text] [doi: 10.1210/jc.2016-2573] [Medline: 28359099]
47. Bhutani VK, Johnson L, Sivieri EM. Predictive ability of a predischarge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near-term newborns. Pediatrics 1999 Jan;103(1):6-14. [doi: 10.1542/peds.103.1.6] [Medline: 9917432]
48. American Academy of Pediatrics Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. Pediatrics 2004 Jul;114(1):297-316. [doi: 10.1542/peds.114.1.297] [Medline: 15231951]
49. Goldberg DL, Becker PJ, Brigham K, Carlson S, Fleck L, Gollins L, et al. Identifying malnutrition in preterm and neonatal populations: recommended indicators. J Acad Nutr Diet 2018 Sep;118(9):1571-1582. [doi: 10.1016/j.jand.2017.10.006] [Medline: 29398569]
50. Ramdin T, Ballot D, Rakotsoane D, Madzudzo L, Brown N, Chirwa T, et al. Neurodevelopmental outcome of late preterm infants in Johannesburg, South Africa. BMC Pediatr 2018 Oct 15;18(1):326 [FREE Full text] [doi: 10.1186/s12887-018-1296-3] [Medline: 30322374]
51. Pratt JS, Browne A, Browne NT, Bruzoni M, Cohen M, Desai A, et al. ASMBS pediatric metabolic and bariatric surgery guidelines, 2018. Surg Obes Relat Dis 2018 Jul;14(7):882-901 [FREE Full text] [doi: 10.1016/j.soard.2018.03.019] [Medline: 30077361]
52. Ekvall S, Ekvall VK. Pediatric Nutrition in Chronic Diseases and Developmental Disorders: Prevention, Assessment, and Treatment. Third Edition. Oxford: Oxford University Press; 2016.
53. Becker P, Carney LN, Corkins MR, Monczka J, Smith E, Smith SE, Academy of Nutrition and Dietetics, American Society for Parenteral and Enteral Nutrition. Consensus statement of the Academy of Nutrition and Dietetics/American Society for Parenteral and Enteral Nutrition: indicators recommended for the identification and documentation of pediatric malnutrition (undernutrition). Nutr Clin Pract 2015 Feb;30(1):147-161. [doi: 10.1177/0884533614557642] [Medline: 25422273]
54. Pridham KF, Limbo RK, Schroeder MM. Guided Participation in Pediatric Nursing Practice: Relationship-Based Teaching and Learning With Parents, Children, and Adolescents. New York: Springer; 2018.
55. Australian Pediatric Endocrine Group. Growth & Growth Charts URL: https://apeg.org.au/clinical-resources-links/growth-growth-charts/ [accessed 2019-11-28]
56. Iacobelli S, Viaud M, Lapillonne A, Robillard P, Gouyon J, Bonsante F, NUTRIQUAL Group. Nutrition practice, compliance to guidelines and postnatal growth in moderately premature babies: the NUTRIQUAL French survey. BMC Pediatr 2015 Sep 3;15:110 [FREE Full text] [doi: 10.1186/s12887-015-0426-4] [Medline: 26337814]
57. Itabashi K, Miura F, Uehara R, Nakamura Y. New Japanese neonatal anthropometric charts for gestational age at birth. Pediatr Int 2014 Oct;56(5):702-708. [doi: 10.1111/ped.12331] [Medline: 24617834]
58. Li H, Capital Institute of Pediatrics, Coordinating Study Group of Nine Cities on the Physical Growth and Development of Children. [Growth standardized values and curves based on weight, length/height and head circumference for Chinese children under 7 years of age]. Zhonghua Er Ke Za Zhi 2009 Mar;47(3):173-178. [Medline: 19573429]
59. Apple App Store. PediTools: Fenton 2013 URL: https://apps.apple.com/us/app/peditools-fenton-2013/id1457464960 [accessed 2019-11-28]
60. Apple App Store. PediTools: GA Calc URL: https://apps.apple.com/us/app/peditools-ga-calc/id1474552213 [accessed 2019-11-28]
61. Gauld LM, Kappers J, Carlin JB, Robertson CF. Height prediction from ulna length. Dev Med Child Neurol 2004 Jul;46(7):475-480 [FREE Full text] [doi: 10.1017/s0012162204000787] [Medline: 15230461]
62. Ohuma EO, Altman DG, International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st Project). Statistical methodology for constructing gestational age-related charts using cross-sectional and longitudinal data: The INTERGROWTH-21 project as a case study. Stat Med 2019 Aug 30;38(19):3507-3526 [FREE Full text] [doi: 10.1002/sim.8018] [Medline: 30488491]
63. Flaherman VJ, Schaefer EW, Kuzniewicz MW, Li SX, Walsh EM, Paul IM. Early weight loss nomograms for exclusively breastfed newborns. Pediatrics 2015 Jan;135(1):e16-e23 [FREE Full text] [doi: 10.1542/peds.2014-1532] [Medline: 25554815]
64. Villar J, Ismail LC, Victora CG, Ohuma EO, Bertino E, Altman DG, International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. Lancet 2014 Sep 6;384(9946):857-868. [doi: 10.1016/S0140-6736(14)60932-6] [Medline: 25209487]
65. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). J R Stat Soc Ser C Appl Stat 2005 Jun;54(3):507-554 [FREE Full text] [doi: 10.1111/j.1467-9876.2005.00510.x]

66. Horbar JD, Ehrenkranz RA, Badger GJ, Edwards EM, Morrow KA, Soll RF, et al. Weight growth velocity and postnatal growth failure in infants 501 to 1500 grams: 2000-2013. Pediatrics 2015 Jul;136(1):e84-e92 [FREE Full text] [doi: 10.1542/peds.2015-0129] [Medline: 26101360]
67. Griffin IJ, Tancredi DJ, Bertino E, Lee HC, Profit J. Postnatal growth failure in very low birthweight infants born between 2005 and 2012. Arch Dis Child Fetal Neonatal Ed 2016 Jan;101(1):F50-F55. [doi: 10.1136/archdischild-2014-308095] [Medline: 26201534]

Abbreviations

CDC: Centers for Disease Control and Prevention
LMS: lambda-mu-sigma
MIT: Massachusetts Institute of Technology
NICU: neonatal intensive care unit
VON: Vermont Oxford Network
WHO: World Health Organization
ΔZ: change in weight Z-score
ΔZ weight: change in weight Z-score from birth to discharge

Edited by G Eysenbach; submitted 12.09.19; peer-reviewed by P Becker, D Goldberg, T Fenton; comments to author 22.10.19; revised version received 30.10.19; accepted 13.11.19; published 30.01.20.

Please cite as:
Chou JH, Roumiantsev S, Singh R
PediTools Electronic Growth Chart Calculators: Applications in Clinical Care, Research, and Quality Improvement
J Med Internet Res 2020;22(1):e16204
URL: https://www.jmir.org/2020/1/e16204
doi:10.2196/16204
PMID:32012066

Original Paper

An Online Pain Education Program for Working Adults: Pilot Randomized Controlled Trial

Yajie Li[1], MSc; Mun Yee Mimi Tse[1], PhD

School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Corresponding Author:
Mun Yee Mimi Tse, PhD
School of Nursing
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Phone: 852 2766 6541
Email: mimi.tse@polyu.edu.hk

Abstract

Background: Pain is a common public health concern, and the pain situation among the general population is serious in mainland China. Working adults commonly experience pain because of long sitting times, a lack of free time, and exercise. A lack of pain-related knowledge is also a significant factor. Educational and therapeutic programs delivered online were used more often in Western countries, and accessible programs in China are limited, especially for pain management. Therefore, we carried out an online pain education program for working adults to self-manage pain. The program was delivered through WeChat, a popular and secure social media with a large population base in China.

Objective: This study aimed to (1) provide pain-related knowledge and self-relief strategies, (2) help participants reduce pain and improve pain-related emotional well-being, and (3) explore participants' learning performance and the acceptability of the online pain education program.

Methods: This was a randomized controlled trial. Chinese adults aged between 16 and 60 years with full-time employment, with pain in the past 6 months, and without any mental illness were recruited using snowball sampling through the internet and were randomly allocated to an experimental group and a control group in 1:1 ratio after the baseline assessment. The 4-week educational program that included basic knowledge of pain, pharmacological and nonpharmacological treatments, and related resources was provided only to the experimental group. Outcomes of pain, depression, anxiety, stress, and pain self-efficacy were measured at baseline (T0), posttreatment (T1), and 1-month follow-up (T2). Participants' acceptability and satisfaction were explored after completing the educational program.

Results: In total, 95 eligible participants joined in the program: 47 in the experimental group and 48 in the control group. Neck and shoulder, head, and back were most commonly reported pain sites with high pain scores. Pain intensity and interference of the experimental group were significantly reduced after the educational program. Depression, anxiety, and stress clinically improved and pain self-efficacy improved after the educational program. The difference in depression, anxiety, stress, and pain self-efficacy within a group or between groups was not statistically significant; however, clinical improvements were demonstrated. A significant correlation between dosage of the intervention and pain intensity and depression was demonstrated. After completing the educational program, more than half of the participants showed acceptance of and satisfaction with the program, and they were willing to recommend the program to others.

Conclusions: Our findings highlight the significant potential of this online education program in the treatment of pain.

Trial Registration: ClinicalTrials.gov NCT03952910; https://clinicaltrials.gov/ct2/show/NCT03952910

(J Med Internet Res 2020;22(1):e15071) doi:10.2196/15071

KEYWORDS

pain; online education; WeChat; working adults

Introduction

Pain is a common and major public health concern [1-3] with a high negative impact on different aspects of the affected individual's quality of life [4-8]. Pain prevalence in developing countries was reported to be approximately 40% among the general population [9]. In Asia, the prevalence of pain in adults ranges from 7.1% to 61% [10]. Studies from China showed a serious pain situation where the estimation of pain prevalence was approxiamately 40% [11-13]. The pain situation of working populations should be taken into consideration, as long sitting time and computer-facing time can lead to discomfort of the body, especially in the neck, shoulders, and back [14].

Many face-to-face pain management programs have been carried out to control pain and reduce its negative impact [5,15-17]. However, the internet has been used as an innovative approach to deliver these programs using the same principles, providing same evidence-based treatments, and teaching the same skills as those delivered face to face [1,18,19]. The internet offers a viable way to deliver self-management support for assisting patients in managing a wide variety of conditions and has the potential to overcome many barriers of the face-to-face approach. One of the obvious benefits is availability of the programs; participants can access them at their convenience and pace, which may provide better control of their situation and yield a greater outcome [20]. As the use of the internet and social networking increased, the increase in health care use via these modes was inevitable [21]. Increasing evidence shows that internet-delivered educational and therapeutic treatments have high accessibility and acceptability [22].

The internet is widely used in China. The China Internet Network Information Center reported that more than 55% of the Chinese population was using the internet by December 2017; among these internet users, 97.5% were using a mobile phone [23]. WeChat is a popular free mobile app for communication and accessing the internet. WeChat attracted more than 900 million active users as of September 2017 [24]. In recent years, subscription, as a new plug-in in WeChat, is a new means to propagate information under a safe condition and is becoming increasingly popular [25,26].

Although the use of online programs to help people with pain is a logical way to overcome many existing barriers, there is limited research in China focusing on illustrating the effectiveness of such programs among working adults. To the best of our knowledge, only one study used the internet to deliver a pain management program for teenage girls to self-manage dysmenorrhea in China and proved its effectiveness [27]. Therefore, we conducted an online pain education program through WeChat for self-management of pain among working adults. The study evaluated the effectiveness of the online pain education program in reducing pain and improving pain-related emotional well-being. We also evaluated participants' learning performance and acceptability.

Methods

Study Design

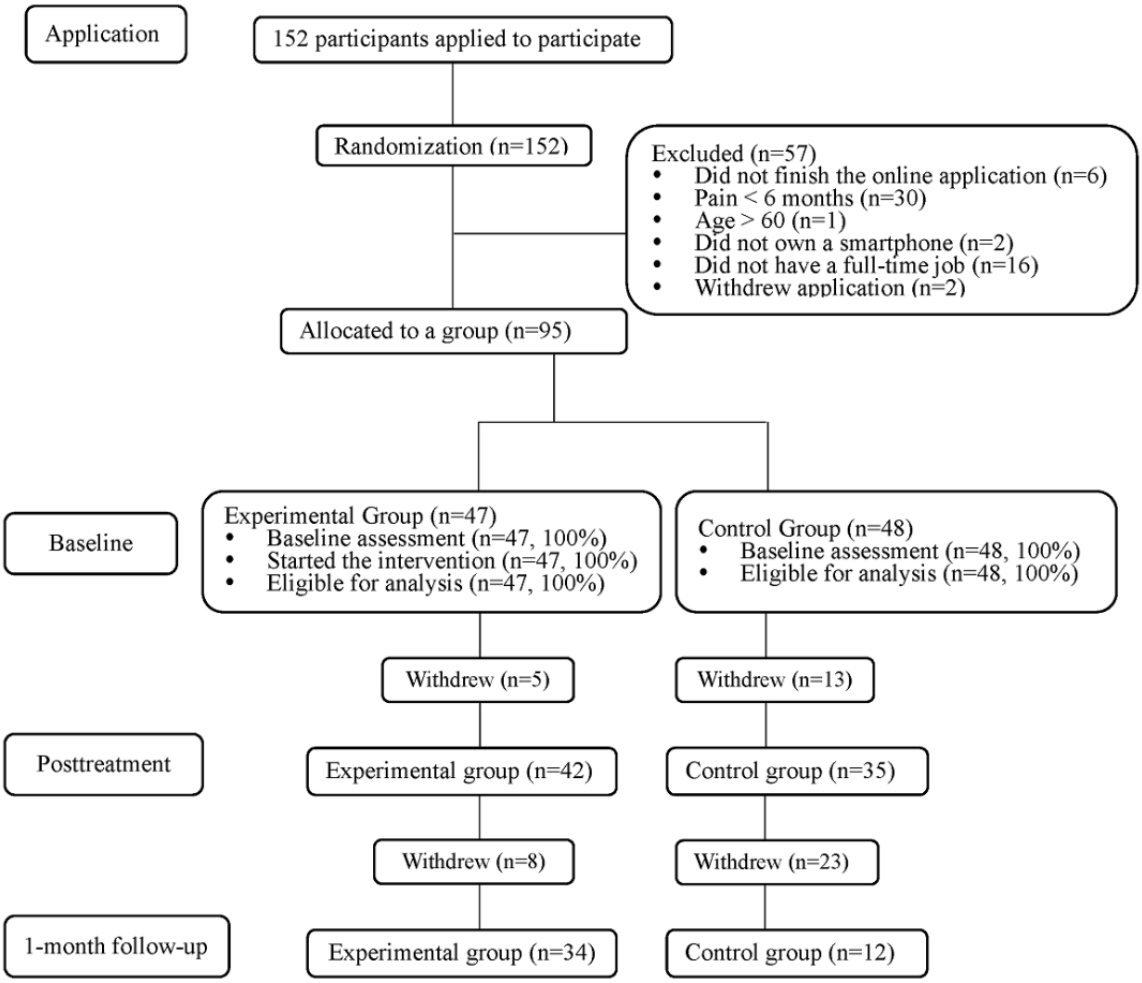
The study was designed as a randomized controlled trial that examined the effectiveness of a 4-week online pain education program. The study was approved by the ethical committee of The Hong Kong Polytechnic University (ref. HSEARS20180519002). Data were collected from September 2018 to March 2019.

Participants

An online poster explaining the details with a quick response (QR) code for the app was designed and distributed in WeChat to attract participants. Individuals who were interested in participating could scan the QR code to register in the study. A total of 152 people applied to join in the program. The application process involved completing an online questionnaire to screen eligibility. Eligible participants were then randomly allocated to one of the two groups: (1) the group that received the pain education program (experimental group) or (2) the group that received only simple material (control group). To minimize the potential of study bias, randomization was performed using an online randomizer [28] with 1:1 ratio after the registration period by a person who was not involved in this study.

Participants were required to fulfill the following criteria: (1) presence of noncancer pain in the past 6 months, with a pain score of at least 2 when assessed using an 11-point scale; (2) age between 16 and 60 years; (3) full-time employment; (4) ability to understand Chinese; and (5) ownership of a smartphone to access the internet. Those who had mental disorders, drug addiction problem, or further treatments planned were excluded from this study. Of the 152 people registered, 95 fulfilled the criteria: 47 were allocated to the experimental group, and 48 were allocated to the control group. The Consolidated Standards of Reporting Trials map for this study is shown in Figure 1.

Figure 1. Consolidated Standards of Reporting Trials map.



Ethics, Consent, and Permissions

Before the eligibility assessment, individuals' consent was obtained electronically, emphasizing that participation was voluntary and remunerative. It was clarified that withdrawal from the study was accepted at any stage. Participants were also informed that all their personal information would remain confidential.

Experimental Group Versus Control Group

Experimental Group

The intervention provided for the experimental group in this study was the online pain education program, which encouraged participants to learn and practice the knowledge and skills introduced in the program. An overview of the program is presented in Multimedia Appendix 1.

The program content including basic knowledge of pain, physical and psychological impact of pain, pharmacological and nonpharmacological treatments, and relevant resources were uploaded to the subscription in a short article format at the beginning of the program. Each article took approximately 3-5 min to read without order restriction. In addition, participants could interact with each other in the program. At the end of each article, three to four multiple-choice questions (MCQs) were asked depending on the article, and all the participants were required to answer the questions. Unlimited attempts were allowed for the MCQs, and log in with WeChat ID was required. Correct answers were provided and would be available after completing the MCQs. The score of each participant was used to evaluate an individual's learning performance. In addition, a WeChat group was created for the participants in the experimental group, and they were encouraged to discuss issues related to the pain education program and share their learning experience. Materials were always accessible during the available period. Regular reminders on a weekly basis were sent to the participants through WeChat.

A total of five experts were invited to assess the content validity using content validity index, including two registered nurses, two pain specialists, and one expert in traditional Chinese medicine. The result of the content validity index was 0.95, which indicated that the program was validated [29]. Test-retest reliability was performed by 10 people 2 weeks apart, and the results ranged from 0.82 to 0.96, suggesting that the content was reliable [30].

Control Group

Brief (one page) material related to pain (that was obtained from an online leaflet open for public) from a grade A tertiary hospital in China [31] was given to the control group in WeChat at the beginning of the program. These participants were required to

read the material whenever possible during the 4-week study period.

Outcome Measures

Outcome measures were administered online at three time points: (1) baseline (T0): after randomization and before starting the education program, (2) posttreatment (T1): right after the experimental group finished the program, and (3) follow-up (T2): 1 month after finishing the program. A battery of well-designed questionnaires was used for outcome measures. The primary and secondary measures were administered at T0, T1, and T2. Participants' learning performance, satisfaction, and acceptability assessment were administered after the intervention (ie, T1). The questionnaires were uploaded in WeChat at three outcome measurement points. To facilitate a high completion rate of the assessment, a reminder message was sent to the participants individually.

Primary Outcome

Brief Pain Inventory - Chinese Version

The Brief Pain Inventory (BPI) is a brief, self-administered questionnaire, which is designed to measure the pain intensity and impairment caused by pain. It consists of four questions related to pain severity and seven questions related to pain interference. The pain interference items focus on general activities, mood, walking ability, work, relationship with others, sleep, and enjoyment of life. A previous study showed that the BPI can be used to measure cancer pain as well as chronic pain and proved that the Chinese version of BPI (BPI-C) has good internal consistency and acceptable test-retest reliability [32].

Secondary Outcome

Depression Anxiety Stress Scales-21 - Chinese Version

Depression Anxiety Stress Scales-21 (DASS-21) is a self-report instrument to measure three negative emotional states: depression, anxiety, and stress. A higher score indicates a greater level of psychological symptoms. A previous study demonstrated that the Chinese version of DASS-21 has excellent internal consistency and validity [33].

Pain Self-Efficacy Questionnaire - Chinese Version

The Pain Self-Efficacy Questionnaire - Chinese Version (PSEQ-C) contains 10 questions regarding a patient's belief about his or her ability to accomplish the daily tasks despite pain. A higher score reflects stronger pain-related self-efficacy. Internal consistency and validity of the PSEQ-C have been proved [34].

Satisfaction, Acceptability, and Learning Performance

Satisfaction and acceptability measures were assessed at the posttreatment assessment. Several questions were asked at the end of the program to assess participants' satisfaction with and acceptability of the program, such as (1) "Do you think the program is useful?" (2) "Does it worth your time?" and (3) "Would you feel confident to recommend this program?" The questions were used in previous studies to assess the acceptability of the internet-delivered program [19,35]. Open-ended questions were also used, including "How do you think about this program?" and "What are the strengths/disadvantages of the program?" [36] Learning performance was measured at T1 and T2. The score of the MCQs was calculated. A total score of ≥10 was considered a better learning performance.

Statistical Analysis

The Statistical Package for the Social Sciences (SPSS) version 23 (IBM corporation, Armonk, New York) was used for handling and analyzing the data. The outcome variables and demographic characteristics were presented using descriptive statistics. The differences in demographic characteristics and outcome variables between the two groups were compared using a Chi-square test. Independent sample *t* test was applied to compare the changes in mean scores of the outcome variables. One-way analysis of variance was conducted to test the within-group changes of the outcome at baseline, posttreatment, and 1-month follow-up. Bivariate correlation was used to assess the correlation between the dosage of intervention (ie, frequency of reading the online materials in WeChat) and the outcome variables. The significance level was set at .05 (two tailed); a *P* value <.05 was considered statistically significant. Responses to open-ended questions on satisfaction with and acceptability of this online program were analyzed using a conventional content analysis.

Results

Baseline Characteristics

The baseline demographics, pain-related characteristics, and baseline outcome of all the participants are presented in Multimedia Appendix 2. The results suggested that more female participants experienced pain than male participants in both groups. More than half the participants were aged between 21 and 30 years. The study involved a predominantly college-educated population (91/95, 96%). In all, 34.7% of the participants were professionals, and 20% had a monthly salary over 10,000 CNY (US $1488), which accounts for the highest proportion in our study. Most of the participants were living in the Southern and Northwest China. There was no significant difference between any of the baseline characteristics of the two groups.

Pain: Experimental Group Versus Control Group Over Time

Pain Intensity and Pain Interference

As presented in Table 1, the overall mean pain score of the experimental group was significantly lower than that of the control group (*P*=.001). Pain intensity of the experimental group was also significantly different between baseline and posttreatment, while no such difference was observed in the control group. Pain interference improved in the experimental group at T1, and the within-group difference in the experimental group showed statistical significance (*P*<.01). The between-group difference was also statistically significant.

Table 1. Pain: Experimental group versus control group over time.

Group (time point)	Experimental group, mean (SD)	Control group, mean (SD)	Mean difference	Cohen *d*[a] (95% CI)	*P* value[b]
Pain intensity					
Baseline (T0)	4.19 (2.07)[c]	4.02 (2.19)	0.171	0.080 (−0.698 to 1.039)	.70
Posttreatment (T1)	3.17 (1.15)[c]	4.26 (1.60)	−1.090	−0.784 (−1.715 to −0.466)	.001
One-month follow-up (T2)	3.85 (1.58)	3.58 (2.07)	0.270	0.147 (−0.890 to 1.429)	.64
Pain interference					
Baseline (T0)	2.75 (1.53)[d]	2.84 (1.40)	–0.086	−0.051 (−0.768 to 0.597)	.80
Posttreatment (T1)	2.36 (0.40)[d]	2.98 (0.67)	–0.620	−1.139 (−0.872 to −0.381)	<.001
One-month follow-up (T2)	3.11 (1.89)	2.71 (1.10)	0.400	0.255 (−0.772 to 1.559)	.50
Pain self-efficacy					
Baseline (T0)	43.09 (15.46)	46.38 (14.43)	−3.290	−0.220 (−9.382 to 2.803)	.29
Posttreatment (T1)	46.52 (8.83)	45.34 (10.04)	1.181	0.125 (−3.103 to 5.465)	.58
One-month follow-up (T2)	46.12 (11.44)	47.25 (11.03)	−1.132	−0.101(−8.806 to 6.541)	.77

[a]Guideline for Cohen *d*: small, *d*=0.2; medium, *d*=0.5; and large, *d*=0.8.

[b]Independent sample *t* test was applied. A *P* value <.05 was considered statistically significant.

[c]One-way analysis of variance was applied, *P*=.012. Pain intensity at T0 was greater than that at T1.

[d]One-way analysis of variance was applied, *P*<.01. Pain interference at T0 was greater than that at T1.

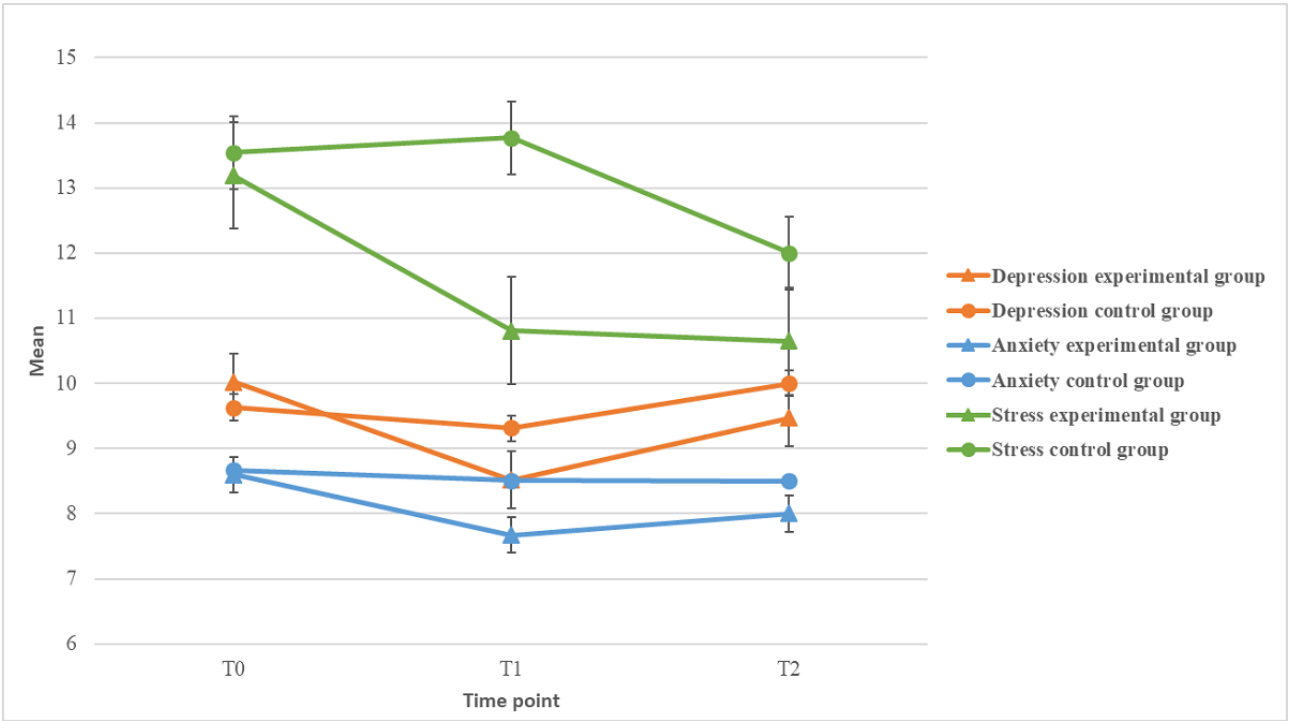
Pain Self-Efficacy

Results of pain self-efficacy questionnaire in the two groups over time are reported in Table 1. A clinical improvement in pain self-efficacy was observed in the experimental group after the program. There were no statistically significant changes in the control group. The between-group differences were not significant over time.

Depression, Anxiety, and Stress

Changes in depression, anxiety, and stress are shown in Figure 2. Improvements in depression, anxiety, and stress were shown in the experimental group; however, the differences were not significant. Depression and anxiety in the control group reduced slightly, whereas the stress level increased slightly. The between-group differences were nonsignificant.

Figure 2. Depression, anxiety, and stress over time.



Dosage of Intervention (Frequency of Reading the Online Materials in WeChat)

The average dosage of intervention (ie, frequency of reading the online materials in WeChat) in the education period (ie, T0 to T1) was 4.07 (SD 1.76), which is higher than that in the follow-up period (ie, T1 to T2; 2.68 [1.72]). The difference between the two periods was statistically significant (P=.001). The correlation between the dosage of intervention and pain intensity, depression, anxiety, and stress was demonstrated. At the 1-month follow-up, a significant correlation was observed between the dosage of intervention and pain intensity/depression (Table 2).

Table 2. Correlation between dosage and outcome variables.

Variable	T1 (posttreatment)		T2 (1-month follow-up)	
	r [a]	P value	r	P value
Pain intensity	−0.393[b]	.01	−0.599[b]	<.001
Pain interference	0.217	.17	−0.180	.31
Pain self-efficacy	0.081	.61	0.122	.49
Depression	−0.564[b]	<.001	−0.726[b]	<.001
Anxiety	−0.316[c]	.04	−0.070	.70
Stress	−0.310[c]	.05	−0.040	.82

[a]r is calculated using the Pearson correlation. Guideline: small, r=0.10 to 0.29; medium, r=0.30 to 0.49; large, r=0.50 to 1.0.

[b]Correlation is significant at .01 level (two-tailed).

[c]Correlation is significant at .05 level (two-tailed).

Learning Performance, Satisfaction, and Acceptability of the Online Program

The overall mean score of the MCQs for the experimental group was 9.67 (SD 1.028) of 10. In all, 76% (32/42) participants reported that they were satisfied with this online pain education program, and 69% (29/42) felt it was worth spending time on. Moreover, 33 (78.57%) participants showed willingness to recommend this program to others. Answers for open-ended questions also showed participants' satisfaction and acceptability: "the program is quite convenient," "the knowledge is useful," and "will recommend to others."

Discussion

Principal Findings

This study aimed to evaluate the effectiveness of an online pain education program and participants' satisfaction and acceptability. After our education program, the pain intensity reduced significantly in the experimental group, and depression, anxiety, stress, and pain self-efficacy showed clinical improvement. A significant correlation was demonstrated between depression, anxiety, and dosage of the intervention. Our results also showed that this online program was acceptable, and participants were satisfied with the program and willing to recommend it to others.

In this study, participants were recruited online via WeChat. Open recruitment through local media or the internet is likely to attract individuals who are more motivated to participate [37]. Consistent with this finding, we found a high completion rate of 89% (42/27) in the experimental group. Results showed that the learning performance of participants was acceptable with the mean score of 9.67/10. This may be because of the socioeconomic status of the participants. Research has indicated that the proportion of people with reliable access to the internet is lower among those with lower socioeconomic status [18,38,39], and a higher risk of pain is correlated with a lower sociodemographic status [11,18,40]. Although the emerging evidence of the internet-delivered pain management program is encouraging, there are still many people who do not have reliable access to the internet and thus have difficulty in utilizing the online programs [18]. As reported by the National Bureau of Statistics of China in 2017, the average monthly income of general Chinese population was 2165 Yuan (US $318) and was lower in the rural areas [41]. The monthly income of approximately 73% of our participants was higher than the average among the general population. The income was high enough to cover the cost of the internet. In addition, participants with a high educational level accounted for the highest proportion. The good learning ability is indicated with the high level of education. Therefore, for most of the working adults in China, sufficient affordability and learning ability make the online pain education program applicable.

Major findings according to the primary outcome measures suggested that this pain education program has the potential to reduce the pain intensity. The result of the significant reduction of pain intensity is consistent with the previous studies conducted among different population in other countries [4,16,18,42]. Studies on internet-delivered pain management programs demonstrated that participants' symptoms are relatively stable over time among those in the treatment-as-usual control group (without a target intervention) [18,19,36,43].

In our study, the significant reduction in pain interference indicated that participants had become more functional and could self-manage their pain more effectively in the daily life [44]. In the experimental group, compared with the control group, the reduction in pain interference was significant, indicating that our online pain education program has the

potential to reduce disturbances caused by pain. Our results are consistent with those of previous online pain management studies that reported improved pain interference after the intervention [44-46].

However, the changes in pain intensity and pain interference were statistically nonsignificant in the control group, and the differences in the 1-month follow-up in this study were not significant between the two groups. This may relate to the loss to follow-up rate. As noted in this study, a rate higher than 50% (23/35) was demonstrated in the control group, which is considered one of the reasons for the nonsignificant differences within the control group over time. In addition, the one-page simple material provided at beginning to the control group perhaps was insufficient, and no interaction with participants may impact the outcome as well.

Only clinical improvements were observed in depression, anxiety, stress, and self-efficacy because of the length of the intervention and the follow-up period; a period of 1 month is possibly too short to achieve a significant effect. Contrary to the expectations, the stress level of the control group increased at the posttreatment assessment. This unexpected result may be because of the factors not related to the education program, such as participants' heavy and busy work, as the participants were working adults. Other stress from their daily life rather than pain-related stress may also have impacted the outcome.

It is noteworthy that a significant correlation between the dosage of the intervention and the outcome variables was demonstrated in our study. In a pain self-management study, Nicholas et al [15] reported that a higher dose of the intervention could partially explain the better outcome achieved, which was consistent with our findings that a higher dosage of intervention resulted in less pain and better pain-related emotional well-being. However, the previous study provided the intervention using a face-to-face approach. Our results proved that dosage was an important factor that impacts the outcome in the internet-delivered program, which is similar to the face-to-face program.

The higher dosage of intervention in the education period compared with the follow-up is probably because of the reminders we sent. The reminders were sent regularly on a weekly basis, with the purpose of encouraging and supporting the participants to work through the education program and keep reading the materials. Although we expected the participants to read the materials and learn consciously, sometimes, participants may forget to do so when they are busy with work. A regular reminder is a good way to enhance participation. In the previous online pain management programs, emails or phone calls were used to prompt individuals to learn the information and apply the skills taught, and they also helped decrease the dropout rate and facilitate compliance [4,18,36,47]. An internet-based intervention study successfully used such email prompts to encourage use and return visits to online resources [48]. In addition, it was demonstrated that such prompts do not compromise the clinical outcomes and acceptability of the program [49]. WeChat massages were used as an alternative in our study and were effective in enhancing participation.

A previous online pain management program stated it did not involve any interactive component [19]; thus, the interactive module designed in our study was unique. Participants' understanding of the knowledge provided can be improved during the interaction. The MCQs designed after each article can encourage participants to apply the knowledge learned immediately. Involving some interaction in an education program, particularly for the online program, is necessary. A previous study highlighted the importance of interaction in high-quality education [50]; the researchers stated concluded that the interaction is the most fundamental form of the education. In addition, increasing the amount of interaction can lead to more effective learning and improve satisfaction with the education program as well as the learning outcome [51]. Wright [52] illustrated that if the intervention includes a forum where participants can interact and support each other, they can gain some benefits from similar experiences of other participants. The interactive component including the MCQs and WeChat group designed in this study played a crucial role.

Our study has a number of strengths. First, this study was the first randomized controlled trial designed to explore an education program using an online approach for working adults to self-manage pain in China. Second, a high completion rate of the education program and questionnaires was achieved. Third, an interactive component was involved in this online program, and participants' learning performance was evaluated. Fourth, we explored the correlation between the dosage of intervention and outcome. Finally, we supplemented satisfaction and acceptability ratings with qualitative analysis of participants' feedback.

Limitations

There are several limitations to this study. First, the sample size of the study was relatively small, which limited the statistical power to detect smaller effects. It is possible that significant differences would have been observed with a large sample size. In addition, the 1-month follow-up of participants in our study may not be sufficient. Indeed, a long-term follow-up (eg, ≥6 months) would assist in observing the long-term effect of the program [36]. The completion rate of 48.4% in the 1-month follow-up may indicate the need to have more frequent reminders, to retain the participants in the program in future studies.

Conclusions

Our findings highlight the significant potential of this online education program in the treatment of pain. Pain intensity reduced significantly after the education program, and pain-related emotional well-being was found to clinically improve. A significant correlation was demonstrated between depression, anxiety, and dosage of the intervention. We conclude that this online program is acceptable. Further promotion to the public can be made to help more people with pain.

Acknowledgments

We are grateful to all the professionals, families, and friends who contributed to this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Content of the pain education program.
[DOCX File , 16 KB - jmir_v22i1e15071_app1.docx]

Multimedia Appendix 2
Baseline characteristics of participants.
[DOCX File , 19 KB - jmir_v22i1e15071_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[DOC File , 217 KB - jmir_v22i1e15071_app3.doc]

References

1. Hayes S, Hogan M, Dowd H, Doherty E, O'Higgins S, Gabhainn SN, et al. Comparing the clinical-effectiveness and cost-effectiveness of an internet-delivered Acceptance and Commitment Therapy (ACT) intervention with a waiting list control among adults with chronic pain: study protocol for a randomised controlled trial. BMJ Open 2014 Jul 2;4(7):e005092 [FREE Full text] [doi: 10.1136/bmjopen-2014-005092] [Medline: 24993763]
2. Peng WL, Wu GJ, Sun WZ, Chen JC, Huang AT. Multidisciplinary management of cancer pain: a longitudinal retrospective study on a cohort of end-stage cancer patients. J Pain Symptom Manage 2006 Nov;32(5):444-452. [doi: 10.1016/j.jpainsymman.2006.05.017] [Medline: 17085270]
3. Reyes-Gibby CC, Duc NB, Yen NP, Nga NH, Tran TV, Guo H, et al. Status of cancer pain in Hanoi, Vietnam: a hospital-wide survey in a tertiary cancer treatment center. J Pain Symptom Manage 2006 May;31(5):431-439. [doi: 10.1016/j.jpainsymman.2005.10.005] [Medline: 16716873]
4. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
5. Morone NE, Greco CM, Weiner DK. Mindfulness meditation for the treatment of chronic low back pain in older adults: a randomized controlled pilot study. Pain 2008 Feb;134(3):310-319 [FREE Full text] [doi: 10.1016/j.pain.2007.04.038] [Medline: 17544212]
6. Ausmed. 2016 Feb 2. The Emotional and Psychological Impacts of Chronic Pain URL: https://www.ausmed.com/articles/chronic-pain-emotional-psychological-impacts/ [accessed 2019-11-15]
7. de Vlieger P, Crombez G, Eccleston C. Worrying about chronic pain. An examination of worry and problem solving in adults who identify as chronic pain sufferers. Pain 2006 Jan;120(1-2):138-144. [doi: 10.1016/j.pain.2005.10.022] [Medline: 16360276]
8. Hanssen DJ, Naarding P, Collard RM, Comijs HC, Voshaar RC. Physical, lifestyle, psychological, and social determinants of pain intensity, pain disability, and the number of pain locations in depressed older adults. Pain 2014 Oct;155(10):2088-2096. [doi: 10.1016/j.pain.2014.07.019] [Medline: 25072890]
9. Tsang A, von Korff M, Lee S, Alonso J, Karam E, Angermeyer MC, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. J Pain 2008 Oct;9(10):883-891. [doi: 10.1016/j.jpain.2008.05.005] [Medline: 18602869]
10. Mohamed Zaki LR, Hairi NN. A systematic review of the prevalence and measurement of chronic pain in Asian adults. Pain Manag Nurs 2015 Jun;16(3):440-452. [doi: 10.1016/j.pmn.2014.08.012] [Medline: 25439125]
11. Chen B, Li L, Donovan C, Gao Y, Ali G, Jiang Y, et al. Prevalence and characteristics of chronic body pain in China: a national study. Springerplus 2016;5(1):938 [FREE Full text] [doi: 10.1186/s40064-016-2581-y] [Medline: 27386382]
12. Jackson T, Chen H, Iezzi T, Yee M, Chen F. Prevalence and correlates of chronic pain in a random population study of adults in Chongqing, China. Clin J Pain 2014 Apr;30(4):346-352. [doi: 10.1097/AJP.0b013e31829ea1e3] [Medline: 23887340]
13. Li Y, Tse MM. Pain situations among working adults and the educational needs identified: an exploratory survey via WeChat. BMC Public Health 2019 Aug 22;19(1):1149 [FREE Full text] [doi: 10.1186/s12889-019-7503-9] [Medline: 31438930]

14. Ariëns GA, van Mechelen W, Bongers P, Bouter L, van der Wal G. Psychosocial risk factors for neck pain: a systematic review. Am J Ind Med 2001 Feb;39(2):180-193. [doi: 10.1002/1097-0274(200102)39:2<180::aid-ajim1005>3.0.co;2-#] [Medline: 11170160]
15. Nicholas MK, Asghari A, Blyth FM, Wood BM, Murray R, McCabe R, et al. Self-management intervention for chronic pain in older adults: a randomised controlled trial. Pain 2013 Jun;154(6):824-835. [doi: 10.1016/j.pain.2013.02.009] [Medline: 23522927]
16. Cosio D, Lin EH. Effects of a pain education program for veterans with chronic, noncancer pain: a pilot study. J Pain Palliat Care Pharmacother 2013 Dec;27(4):340-349. [doi: 10.3109/15360288.2013.846953] [Medline: 24147960]
17. Chou PL, Lin CC. A pain education programme to improve patient satisfaction with cancer pain management: a randomised control trial. J Clin Nurs 2011 Jul;20(13-14):1858-1869. [doi: 10.1111/j.1365-2702.2011.03740.x] [Medline: 21615576]
18. Dear BF, Gandy M, Karin E, Ricciardi T, Fogliati VJ, McDonald S, et al. The pain course: a randomised controlled trial comparing a remote-delivered chronic pain management program when provided in online and workbook formats. Pain 2017 Jul;158(7):1289-1301. [doi: 10.1097/j.pain.0000000000000916] [Medline: 28394850]
19. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
20. Gerber BS. The chronic disease self-management program: extending reach through the internet. Med Care 2006 Nov;44(11):961-963. [doi: 10.1097/01.mlr.0000244923.80923.1e] [Medline: 17063126]
21. 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]
22. Tse MM, Tang A, Budnick A, Ng SS, Yeung SS. Pain and pain management among university students: online survey and web-based education. Cyberpsychol Behav Soc Netw 2017 May;20(5):305-313. [doi: 10.1089/cyber.2016.0580] [Medline: 28498043]
23. CNNIC. 41st Statistical Report on the Internet Development in China URL: https://cnnic.com.cn/
24. WeChat Blog: Chatterbox. 2017. The 2017 WeChat Data Report URL: https://blog.wechat.com/2017/11/09/the-2017-wechat-data-report/ [accessed 2019-11-15]
25. Zheng B, Zhao K. Yong Hu Shi Yong Wei Xin Ding Yue Hao de Yuan Yin Tan Jiu. West Journal 2015:59.
26. WeChat Help Center. URL: https://help.wechat.com/oshelpcenter [accessed 2018-11-26]
27. Yeh M, Chen H, Houng Y, Liao F. A New Websitelf-Management of Dysmenorrhea Online Learning. Beijing Biomedical Engineering 2006;25(1):95-98 [FREE Full text]
28. Random.org - True Random Number Service. URL: https://www.random.org/
29. 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]
30. 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]
31. Li B. Eastern Theater General Hospital – Neurology. 2018. Pain brochure URL: http://chinaneurology.org/2018/05/07/%E7%96%BC%E7%97%9B%E5%AE%A3%E4%BC%A0%E6%89%8B%E5%86%8C/ [accessed 2019-11-18]
32. Wang XS, Mendoza TR, Gao SZ, Cleeland CS. The Chinese version of the Brief Pain Inventory (BPI-C): its development and use in a study of cancer pain. Pain 1996 Oct;67(2-3):407-416. [doi: 10.1016/0304-3959(96)03147-8] [Medline: 8951936]
33. Wen Y, Wu DX, Lv XJ, Li HG, Liu XC, Yang YP, et al. Psychometric properties of the Chinese Short Version of Depression Anxiety and Stress Scale in Chinese adults. Chinese J Public Heal 2012;28(11):1436-1438. [doi: 10.1213/01.ane.0000255731.24092.a5] [Medline: 17377107]
34. Lim HS, Chen PP, Wong TC, Gin T, Wong E, Chan IS, et al. Validation of the Chinese version of pain self-efficacy questionnaire. Anesth Analg 2007 Apr;104(4):918-923. [doi: 10.1213/01.ane.0000255731.24092.a5] [Medline: 17377107]
35. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
36. Friesen LN, Hadjistavropoulos HD, Schneider LH, Alberts NM, Titov N, Dear BF. Examination of an internet-delivered cognitive behavioural pain management course for adults with fibromyalgia: a randomized controlled trial. Pain 2017 Apr;158(4):593-604. [doi: 10.1097/j.pain.0000000000000802] [Medline: 27984490]
37. 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]
38. Wang JY, 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 Mar 3;13(1):e25 [FREE Full text] [doi: 10.2196/jmir.1534] [Medline: 21371989]
39. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]

40. Miljkovic A, Stipcic A, Braš M, Dorđević V, Brajkovic L, Hayward C, et al. Is experimentally induced pain associated with socioeconomic status? Do poor people hurt more? Med Sci Monit 2014 Jul 17;20:1232-1238 [FREE Full text] [doi: 10.12659/MSM.890714] [Medline: 25029965]
41. National BOSOC. National Bureau of Statistics of China. 2017 Residents Income and Expenditure Consumption Internet URL: http://www.stats.gov.cn/tjsj/ndsj/2017/indexeh.htm
42. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012 Feb;153(2):319-330 [FREE Full text] [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
43. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
44. Dysvik E, Kvaløy JT, Stokkeland R, Natvig GK. The effectiveness of a multidisciplinary pain management programme managing chronic pain on pain perceptions, health-related quality of life and stages of change--A non-randomized controlled study. Int J Nurs Stud 2010 Jul;47(7):826-835. [doi: 10.1016/j.ijnurstu.2009.12.001] [Medline: 20036362]
45. Brattberg G. Internet-based rehabilitation for individuals with chronic pain and burnout: a randomized trial. Int J Rehabil Res 2006 Sep;29(3):221-227. [doi: 10.1097/01.mrr.0000210055.17291.f5] [Medline: 16900043]
46. Devineni T, Blanchard EB. A randomized controlled trial of an internet-based treatment for chronic headache. Behav Res Ther 2005 Mar;43(3):277-292. [doi: 10.1016/j.brat.2004.01.008] [Medline: 15680926]
47. Andersson G, Lundström P, Ström L. Internet-based treatment of headache: does telephone contact add anything? Headache 2003 Apr;43(4):353-361. [doi: 10.1046/j.1526-4610.2003.03070.x] [Medline: 12656706]
48. McNeill LH, Viswanath K, Bennett GG, Puleo E, Emmons KM. Feasibility of using a web-based nutrition intervention among residents of multiethnic working-class neighborhoods. Prev Chronic Dis 2007 Jul;4(3):A55 [FREE Full text] [Medline: 17572959]
49. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 8;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
50. Su B, Bonk C, Magjuka R, Liu X, Lee S. The importance of interaction in web-based education: a program-level case study of online MBA courses. J Interact Online Learn 2005;4(1):1-19 [FREE Full text]
51. Zhang S. Are interaction time and psychological interactivity the same thing in the distance learning television classroom? Educ Technol 1994;34(6):58-64 [FREE Full text]
52. Wright K. Social support within an on-line cancer community: an assessment of emotional support, perceptions of advantages and disadvantages, and motives for using the community from a communication perspective. J Appl Commun Res 2002;30(3):195-209. [doi: 10.1080/00909880216586]

Abbreviations

BPI: Brief Pain Inventory
CVI: Content Validity Index
DASS-21: Depression Anxiety Stress Scales-21
MCQ: multiple choice question
PSEQ-C: Pain Self-Efficacy Questionnaire-Chinese Version

Edited by G Eysenbach; submitted 18.06.19; peer-reviewed by B Thompson, A Boyd; comments to author 27.08.19; revised version received 01.09.19; accepted 22.10.19; published 14.01.20.

Please cite as:
Li Y, Tse MYM
An Online Pain Education Program for Working Adults: Pilot Randomized Controlled Trial
J Med Internet Res 2020;22(1):e15071
URL: https://www.jmir.org/2020/1/e15071
doi:10.2196/15071
PMID:31934865

Original Paper

Patient Recommendations to Improve the Implementation of and Engagement With Portals in Acute Care: Hospital-Based Qualitative Study

S Ryan Greysen[1], MD, MHS; Yimdriuska Magan[2], BA, MPH; Jamie Rosenthal[3], BA; Ronald Jacolbia[4], BA, MSN; Andrew D Auerbach[5], MD, MPH; James D Harrison[5], PhD, MPH

[1]Section of Hospital Medicine, University of Pennsylvania, Philadelphia, PA, United States
[2]School of Medicine, University of California, Davis, Davis, CA, United States
[3]School of Medicine, Boston University, Boston, MA, United States
[4]School of Nursing, University of California, San Francisco, San Francisco, CA, United States
[5]Division of Hospital Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
S Ryan Greysen, MD, MHS
Section of Hospital Medicine
University of Pennsylvania
3400 Spruce Street
5 Maloney, Suite 5030
Philadelphia, PA, 19104
United States
Phone: 1 202 664 6084
Email: ryan.greysen@uphs.upenn.edu

Abstract

Background: The inclusion of patient portals into electronic health records in the inpatient setting lags behind progress in the outpatient setting.

Objective: The aim of this study was to understand patient perceptions of using a portal during an episode of acute care and explore patient-perceived barriers and facilitators to portal use during hospitalization.

Methods: We utilized a mixed methods approach to explore patient experiences in using the portal during hospitalization. All patients received a tablet with a brief tutorial, pre- and postuse surveys, and completed in-person semistructured interviews. Qualitative data were coded using thematic analysis to iteratively develop 18 codes that were integrated into 3 themes framed as patient recommendations to hospitals to improve engagement with the portal during acute care. Themes from these qualitative data guided our approach to the analysis of quantitative data.

Results: We enrolled 97 participants: 53 (53/97, 55%) women, 44 (44/97, 45%) nonwhite with an average age of 48 years (19-81 years), and the average length of hospitalization was 6.4 days. A total of 47 participants (47/97, 48%) had an active portal account, 59 participants (59/97, 61%) owned a smartphone, and 79 participants (79/97, 81%) accessed the internet daily. In total, 3 overarching themes emerged from the qualitative analysis of interviews with these patients during their hospital stay: (1) hospitals should provide both access to a device and bring-your-own-device platform to access the portal; (2) hospitals should provide an orientation both on how to use the device and how to use the portal; and (3) hospitals should ensure portal content is up to date and easy to understand.

Conclusions: Patients independently and consistently identified basic needs for device and portal access, education, and usability. Hospitals should prioritize these areas to enable successful implementation of inpatient portals to promote greater patient engagement during acute care.

Trial Registration: ClinicalTrials.gov NCT00102401; https://clinicaltrials.gov/ct2/show/NCT01970852

(J Med Internet Res 2020;22(1):e13337) doi:10.2196/13337

KEYWORDS

patient portals; hospitalization; patient engagement; qualitative research

Introduction

Background

Ten years after the passage of the Health Information Technology for Economic Clinical Health in 2008 and the resulting Meaningful Use Incentive Program, most hospitals have adopted electronic health records (EHRs) [1]. The inclusion of patient portals into EHRs has been slower, but it has also been steadily rising, and it is projected to expand rapidly in the near future. According to Health Information National Trends Survey data, 5% of patients were using portals in 2008 and 17% of patients were using portals by 2013, with projections that adoption will likely increase to at least 40% by 2020 but that it may be as high as 75% [2]. In addition, there is growing evidence that portal use can improve outcomes, including medication adherence and diseases management [3], and increased patient empowerment and satisfaction with health services [4]. These trends have added urgency to the national effort to increase the use and effectiveness of portals in all phases of care; however, to date, most of this has focused on portal use in the outpatient setting [5-8].

Although the literature exploring barriers and facilitators to portal use in the outpatient setting is robust—and many demographic trends, such as age, race or ethnicity, and socioeconomic status, certainly apply across care settings [9-12]—there are also unique challenges to portal engagement in the hospital, which merit specific attention. Accordingly, there has been increased exploration of technology use in the hospital to engage patients in care in the last 5 years. A systematic review of this literature performed in 2013 describes a wide range of technologies from video games to interactive displays, mobile phone messaging, and other multimedia approaches; however, there were no studies of portal use in the hospital [13]. Our prior work explored the use of tablets [14] or smartphones [15] in a limited number of hospitalized patients, but it did not explore portal-specific barriers or facilitators to engagement. A more recent systematic review found 17 studies specifically focused on portal use in the hospital, but these studies largely explored design features and institution-specific prototypes [16]. Many of these used qualitative methods with relatively small samples (n=8-21) proportionate to the task of improving design and function. These studies have been fundamental for developing consensus around key features that inpatient portals should include [17,18]; however, as more hospitals are now adapting existing platforms rather than designing their own, a deeper understanding of patient experiences using existing portal platforms from a large and diverse inpatient sample is needed to speed up the process of successful implementation in acute care.

Objectives

Accordingly, we conducted in-person, bedside interviews with 97 patients who were provided with a tablet and access to a widely used institutional portal (Epic MyChart). Our objectives were to obtain a deeper understanding of patent perceptions of using a portal during an acute episode of care and explore patient-perceived barriers and facilitators to portal use during hospitalization. On the basis of our findings, we developed specific recommendations for hospitals and health systems on how to improve portal implementation to maximize patient engagement.

Methods

Study Design, Participant Enrollment, and Portal Characteristics

This study reported the qualitative data collected from debrief interviews with adult patients hospitalized for a general medical condition at the University of California, San Francisco (UCSF), who participated in a randomized controlled trial (RCT) of a focused educational intervention to increase engagement with the portal (ClinicalTrial.gov identifier NCT02109601). Eligibility criteria included the following: admission to the general medicine service, age 18 years or older, and the ability to communicate in English. Exclusion criteria included the following: admission to the intensive care unit (ICU), cognitive impairment, or isolation precautions. Quantitative outcomes data of the randomized trial are reported in a separate paper [19]; here, we discuss qualitative results from debrief interviews with all patients who completed the trial.

Research assistants (RAs) screened patients via UCSF's EHR (Epic) and obtained written consent from those willing to participate. If a participant did not have an active institutional portal account (UCSF MyChart), the RA assisted with registration and activation of a new account. All enrolled participants received an iPad tablet (iPad 16 GB third generation Model A1430) to use for that day only. Participants were instructed on basic iPad features, including how to use the keyboard, home button, and touch screen. As per our RCT protocol reported previously [20], intervention patients (n=50) received an in-person, bedside tutorial on how to navigate the portal, with specific focus on how to perform key tasks that included viewing their test results, viewing medications, messaging with providers, and scheduling appointments. Control patients (n=47) received assistance only with logging in to the portal as needed, but they did not receive detailed guidance or assistance on using the portal to perform the specific tasks above. The UCSF Committee on Human Research (Institutional Review Board) approved the research protocol.

The portal used in this study (Epic MyChart), as configured by UCSF, had the following characteristics. As with most portals, patients can access only certain content, not all information in EHR. UCSF's MyChart provides patient access to many features of the EMR, which are standard in the MyChart platform, including Allergies, Demographics, Health Goals, Medications, Problems, Immunizations, Care Team, Documents, Health History, Lab Results, Plan of Care, Procedures, and Vitals. Most lab results (eg, common blood tests, such as complete blood count or comprehensive metabolic panels) are available in real time, whereas others (eg, advanced imaging results from computed tomography) are released after a 24-hour delay. More detailed information about the portal can be obtained from the Terms and Conditions published on the UCSF MyChart website [21].

Data Collection

Interviews were conducted by 2 RAs (YM and RJ) who received study-specific training in qualitative interviewing techniques. RAs performed 1 interview with each patient at the time the study iPad was recollected, comprising 10 questions: 4 multiple choice questions about patient satisfaction and 6 open-ended questions about their experience (Multimedia Appendix 1). RAs read all interview questions aloud to patients and manually transcribed their responses into a single, secure website (REDCap) in real time [22]. RAs were trained to probe deeper into patient's initial responses using follow-on queries, such as "Just to be sure I understand, what do you mean by X" or "That's interesting, can you tell me more about Y?" RAs read patient responses back to the patient before finalizing each entry to confirm accuracy. To ensure high-quality data collection, 3 study investigators (JDH, ADA, and SRG) gave weekly feedback to RAs, and the first author (SRG) met with RAs daily to review content and provide assistance and guidance.

Data Analysis

We analyzed qualitative data from open-ended questions using a thematic analysis approach [23]. A total of 2 authors (YM and JR) independently performed primary coding using all of the interview data, and they resolved any discrepancies in the individual codes through negotiation. A third author (SRG) performed secondary coding by reviewing all data and modifying the initial code sheet iteratively as needed to capture all conceptual domains observed in the data. Finally, a fourth author (JDH) reviewed the final code sheet along with data (quotes) to support each code and participated (along with the entire coding team YM, JR, and SRG) in the development of themes through integration of multiple codes into overarching concepts. All study authors reviewed and agreed on the final code structure, which contains 18 codes integrated into 3 overarching themes (Multimedia Appendix 2) framed as patient recommendations to improve portal use in the hospital. We used STATA version 13.1 (College Station) to perform frequency analysis and describe participant characteristics, including age, race, gender, electronic device ownership, frequency of devise use, and frequency of internet use.

Results

We enrolled a diverse sample of 97 hospitalized patients (Table 1). Fifty-three (53/97, 55%) were women, and in terms of race/ethnicity, 44 (44/97, 45%) were nonwhite: 14 (14/97, 14%) black, 9 (9/97, 9%) Asian, and 21 (21/97, 22%) other/declined.

The average age was 48.1 years (range 19-81 years), and the average length of hospitalization was 6.4 days. In terms of previous use of technology, 67 participants (67/97, 69%) reported owning a laptop computer, 57 participants (57/97, 59%) owned a smartphone, 51 participants (51/97, 53%) owned a desktop computer, and 48 participants (48/97, 49%) owned a tablet computer. Only 6 participants (6/97, 6%) did not own any of these devices. In addition, 79 participants (79/97, 81%) had previously looked up health information on the Web, 55 participants (55/97, 57%) had used the internet to communicate with a health care provider, 39 participants (39/97, 41%) had scheduled a medical appointment on the Web, and 34 participants (34/97, 36%) had refilled a prescription for a medication over the internet. With regard to use of the institutional portal (UCSF MyChart) specifically, 52 patients (52/97, 54%) had previous experience (had active accounts) and 45 patients (45/97, 46%) were new users (registered for a new account as part of this study).

Overall patient-reported experience with the tablet and portal was very high: 78 patients (78/97, 80%) were satisfied or very satisfied with using the tablet in the hospital, and 83 patients (83/97, 86%) were satisfied or very satisfied using the tablet to access their portal. Qualitative analysis reinforced this, and most patients offered suggestions about how their experience could be enhanced or expanded to include other patients. We organized these suggestions into 3 overarching and integrating themes: (1) hospitals should provide access to a device and bring-your-own-device (BYOD) platform to access the portal; (2) hospitals should provide an orientation on how to use the device and the portal; and (3) information in the portal should be easy to understand and up to date.

Table 1. Participant characteristics (N=97).

Characteristics	Values, n (%)
Demographics	
Age (years)	
18-49	48 (49)
50-60	41 (42)
≥70	8 (8)
Female, gender	53 (55)
Race or ethnicity	
White	53 (55)
Black	19 (20)
Hispanic	9 (9)
Asian	7 (7)
Other/unknown	9 (9)
Insurance	
Medicaid	23 (24)
Medicare	22 (23)
Private	26 (47)
Self-pay/uninsured	6 (6)
Technology use characteristics	
Own desktop computer	51 (53)
Own laptop computer	67 (69)
Own smartphone	57 (59)
Own tablet computer	48 (49)
Does not own a device	6 (6)
Internet use	
Daily	79 (81)
Several times a week	7 (7)
Once a week or less	6 (6)
Prestudy Web-based health tasks	
Looked up health information	78 (80)
Communicated with provider	55 (57)
Scheduled medical appointment	39 (41)
Refilled prescription	34 (35)
None of these	10 (10)

Recommendation 1: Hospitals Should Provide Access to a Device and Bring-Your-Own-Device Platform to Access the Portal

Overall, the most consistent feedback received was that access to a device and the portal was a very positive experience, and, accordingly, many participants felt strongly that our hospital should strive for this level of engagement as standard of care. When we probed deeper, we discovered there were actually several components worth exploring separately. First, patients recommended the hospital provide devices to every patient who wanted one:

> *It would be nice to give loaners [iPad] out to any patient who wants one.* [50-year-old woman, new portal user]

> *All patients should get a device as opposed to waiting for patients to request one.* [60-year-old woman with previous portal experience]

Furthermore, patients also offered suggestions regarding the use of devices once deployed. The overall concept expressed

most consistently was that hospitals should provide multiple options to increase accessibility of the portal by ensuring that the device itself (or accessories) was adaptable to needs of patients:

> *I wished it [the tablet screen] was not as touch sensitive because sometimes my hands shake and I end up select things without me wanting to.* [38-year-old woman, new portal user]

> *Sometimes my device won't respond because I have callouses from burns lack of circulation in my fingers and these devices work with thermal. So not sure if it is me or the device.* [41-year-old man with previous portal experience]

Patients also suggested ways that accessories or modifications to the device could maximize their ability to engage during their hospital stay. Some patients suggested modifications that could represent changes to both the device and some function of the portal as well (eg, voice recognition and transcription):

> *Provide accessibility programs [such as screen reader] and headphones for patients with poor vision.* [35-year-old woman, new portal user]

> *[It would be nice if the iPads had the ability of transcribing after you speak, such as in Google. This iPad didn't have that ability, is that Siri? Having Siri on MyChart would be very nice!* [47-year-old man with previous portal experience]

Finally, with regard to optimizing patient opportunities to access the portal and engage meaningfully with it during hospital care, patients also suggested the development of a BYOD approach:

> *The iPad is fine, but I like to use my own smartphone. It is actually both a phone and tablet. I'm more comfortable because it's my own and also because it's an android and I feel better using it, I'm more familiar with it. I just feel more comfortable with that.* [41-year-old man with previous portal experience]

> *It seems like this [project] is specific to iPad tablets, but other devices such as androids would also work and in case that patients could bring their own device.* [67-year-old man with previous portal experience]

Recommendation 2: Hospitals Should Provide an Orientation on How to Use the Device and the Portal

Overall, most patients expressed high satisfaction with the orientation they received to the device and the portal. Interestingly, some patients with previous experience articulated the value of reviewing basic use and key functions of the portal to ensure familiarity and competency with these functions before addressing more advanced functions or topics, especially given changes and updates in the portal that occur over time:

> *The tutorial was very helpful for me because I have been trying to get signed up on this thing for a while, but each time I tried, I had issues. And it's hard, because you know I am sick and having to deal with one more thing was just overwhelming. It was nice to have you help me through this process finally.* [38-year-old woman, new portal user]

> *I have been using [the portal] for two years, and I have seen different versions. It seems like with time, it gets more confusing. It is not as intuitive anymore, and you have to guess your way around to accomplish the same tasks.* [42-year-old man with previous portal experience]

In addition to general orientation (or reorientation) related to basic functions, which could be tailored to the participant's level of previous experience, several participants suggested that special modifications be made on the basis of other characteristics, such as age and level of technological savvy or sophistication:

> *It would be a good idea to do a focus group with older folks to see if they like tablets, if they want to use them, if it is easy to use, etc. Some people may not want to participate because they don't have computers.* [67-year-old man with previous portal experience]

> *For someone not as techy, maybe walk patients through a tutorial for those who don't know how to use it. Maybe have a test web site or have a little Q&A.* [38-year-old man, new portal user]

Regardless of age, being technology savvy, or level of prior portal experience, many patients expressed a desire for assistance with device settings to optimize their experience. Often, these were very basic issues, such as how to adjust font size:

> *Maybe at the beginning it would be nice making the font bigger. Maybe bold letters to highlight topics such as test results, or have a button that says: 'can you see this/read this?' and picture of the magnifying glass to make it bigger for older people not as tech savvy…it was hard to read even for me and I have 20/20 vision and I had to make the page larger.* [38-year-old man, new portal user]

> *Some links in MyChart were too close together and it was hard to tap the right choice. Larger font will help to be able to see better. Also, not knowing how to use the iPads, the interface was a mystery.* [48-year-old man, new portal user]

Perhaps, most surprisingly, even patients with previous portal experience and those who felt confident in their ability to use the device and navigate the portal expressed a desire for more assistance with the first steps of access: remembering the Web address for the portal (URL institution-specific portal) and their log-in information (username and password):

> *I have been trying to get signed up on this thing [portal] for a while. My doctor has been telling me about it for a long time, but each time I tried, I had issues. Either it didn't recognize my username or password and it was just difficult.* [38-year-old woman, new portal user]

> *Sometimes confusing if you google "UCSF MyChart," it won't take you to the MyChart page that would allow me to login and I can't remember the right web*

address. [76-year-old man with previous portal experience]

Recommendation 3: Hospitals Should Ensure Portal Content is Up to Date and Easy to Understand

Participants frequently commented on the lack of timely information in their portal. Several participants suggested that the portal would be more useful to them in the hospital if it had more frequently updated information. Others suggested they would prefer to be able to see "everything" in terms of results rather than have access only to results from a limited set of labs, imaging studies, or procedures:

> *There wasn't much [on the portal] but my medications at the moment. I wanted to see more information about my tests; that would be nice.* [42-year-old woman, new portal user]

> *MyChart could have more features and it could be updated quicker: I don't mind a data dump of what physicians see.* [42-year-old man with previous portal experience]

Many participants highlighted challenges they encountered in understanding the content in the portal in terms of medical terminology or "jargon" in their portal. Often, patients expressed a desire to increase both quantity (more data) and quality (more interpretation) of information to enhance the meaning and applicability of information to guide their hospital care:

> *Half of the medical information there [in the portal] is hard to understand. Maybe it's easier for someone with a medical background.* [38-year-old man, new portal user]

> *I don't like the list of your diseases in MyChart because when you click on your disease, it gives very generic information…it's really not that helpful. I want it to be more personalized to my illness.* [61-year-old woman with previous portal experience]

In summary, patient feedback revolved around the experience of being hospitalized and the heightened desire for information in this setting. Accordingly, patients directed suggestions for improvement toward the hospital to increase engagement with their portal during acute care.

Discussion

Principal Findings

This qualitative study of patient portal use during hospitalization is one of the largest, in-depth explorations of the patient experience in a highly diverse sample of inpatients using a widely used platform (Epic MyChart). Previous studies have focused on patient suggestions for design aspects of an ideal portal or policies to promote broad adoption [13-24]; our study builds on this literature to characterize fundamental issues to implementation. Indeed, the most consistent feedback was not about developing advanced new features; patients focused on basic issues, such as providing universal access, orientation, and information. These are issues largely within the control of the hospital, but these may go overlooked, as they seem "simple" and may therefore represent underappreciated barriers to successful implementation. Indeed, many of the issues identified by patients represent foundational issues, which, if not resolved on "day one" of a patient's hospitalization, are likely to prohibit more meaningful, longitudinal use of the portal throughout their hospital stay. There are several ways through which our findings could inform implementation and optimization of patient portals in the hospital.

First and foremost, patients in our study felt strongly that access to a device and support for a BYOD approach were fundamental to ensuring broad and meaningful engagement. Relatively few hospitals have taken the approach of placing a patient-facing device in every room; generally, this has been focused to construction of new hospitals [25,26]. Other institutions have supplied tablets to patients, as needed, in specific units, such as oncology and the Medical ICU [27]. Moreover, applying guidelines for BYOD use in the hospital [28] and providing devices just for patients who do not bring their own may suffice, given that many (if not most) now bring their own devices with them to the hospital [15]. A second step toward universal access could be broader adoption of Application Program Interfaces to integrate more seamlessly with device-specific programs, such as Apple's new Heath Records section, which can securely and automatically interface with the EMR from 40 health systems [29]. Finally, with respect to access, it should always be recalled that even patients who bring their own device may still need additional assistance with device use and portal access in the setting of acute illness and hospitalization. Even patients who ordinarily navigate a touch screen interface may need ease-of-access assistance with a keyboard, mouse, or headphones.

Second, patients in our study felt strongly that hospitals should go beyond access to devices and the portal and ensure all patients are adequately oriented to both the device and portal to facilitate engagement. This finding is in alignment with previous studies of other stakeholders, including clinicians, information systems leaders, and administrators [17,30]; however, there are few that focus on the specific challenge of patient education to leverage technology within the hospital and even fewer that specifically focus on portals [20,31]. Indeed, a recent systematic review by Roberts et al [32] described 9 studies focused on familiarization, training, and ongoing support of technology use during hospitalization, but only 1 study focused on the EHR portal [33]. In this study, the issues identified may seem relatively simple, but they are also foundational; thus, they can represent critical barriers that should be addressed on "day one" of hospitalization. Fortunately, implementation solutions for these issues may be relatively straightforward and present hospitals with opportunities for "quick wins." For example, offering an overview orientation to devices and portals to every patient would likely require some combination of standardized approach (eg, Web-based tutorial), as well as the ability to provide individualized assistance as needed (eg, frontline providers) [34]. Some patient groups may also need approaches specifically tailored to them, such as older [9] and even middle-aged patients [10]. Regardless of age, many patients in this study requested assistance with adjusting features of the tablet, such as font size. Even patients who owned the same tablet sometimes needed assistance to configure the device

for optimal portal use, which aligns with our clinical experience with caring for the hospitalized—they are often not able to accomplish simple self-care tasks that they would otherwise do independently when they are not acutely ill. These issues may be overlooked in many hospitals because of assumptions or inferences about patient experiences and preferences, as well as patients' own hesitancy to ask for help in these areas, if not specifically prompted or offered assistance.

Third, patients in this study recommended that hospitals maximize efforts to ensure the content of portals is up to date and easier to understand, which builds on recent studies. O'Leary et al interviewed 18 hospitalized patients, which emphasized the desire for more information and greater assistance with interpretation [30]. Dalal et al identified challenges about communicating care plans through analysis of messages sent via the portal by 158 hospitalized patients [35]. Similarly, a recent scoping review by Roberts et al found an overarching theme of interactive learning for patients, noting "patients are more accepting of, engaged in, and satisfied with education that is tailored to reflect their personal situation" [29]. Similarly, patients in this study wanted access to more information and wanted it to be delivered more quickly. Findings from this study add further support to recent studies [36,37], suggesting that the default should be to release results automatically, unless specifically requested by the ordering clinician. Although releasing more data directly to patients in a "show me everything" fashion may complicate the challenges of making information in the portal easier to understand, the development of more robust dictionaries with hover or mouse-over functionality, links to high-quality health information sites, and support for self-monitoring programs could help mitigate this problem and improve patients' abilities to engage with their results in real time [38].

This study has several limitations. First, it is a single-site study using 1 EHR portal (Epic MyChart), which may limit generalizability to other systems; however, the Epic platform is the most widely used in the United States, which ensures broad applicability. Second, we provided only 1 device (Apple iPad); we did not provide a variety of devices for patients to choose from or study portal interactions using patient-supplied devices (BYOD). Nonetheless, the themes we present here are relevant to any device a patient might use to access the portal during hospitalization (whether hospital-provided or BYOD); thus, they have broad applicability. Third, as with any qualitative study of patient perspectives, there is potential for participant biases to effect results. To minimize the potential for recall bias, we interviewed patients on the same day, on which they were asked to access their portal; in addition, to minimize social desirability bias, we framed questions to solicit open-ended feedback and avoided close-ended questions, such as whether patients liked or disliked certain features. Fourth, we only enrolled patients who were cognitively intact and could provide feedback on their personal experience with the portal; we did not interview family members or caregivers who are especially important in the care of cognitively impaired patients. Finally, we only gave the participants the opportunity to use the iPad for 1 day, and responses might be different if they were interviewed after having more time to use the device or even after leaving the hospital. Future studies should attempt to address this limitation by following patients longitudinally, to understand how their experiences and needs may vary when transitioning from acute care to postacute care and recovery from hospitalization.

Conclusions

In conclusion, our qualitative findings from a study of a large, diverse sample of hospitalized patients highlight opportunities to improve hospital implementation and patient engagement with the portal care in 3 key areas: access, orientation, and usability. Our findings have important implications for the successful deployment of acute-care patient portals, and they suggest several hospital-level interventions to speed implementation of existing platforms. As patients become increasingly engaged with mobile and connected devices in their personal lives, expectations for the use of these technologies to facilitate better engagement during hospitalization will continue to grow. Optimization of their experience via the patient portal is a first and critical step toward realizing the potential for these technologies to improve outcomes in inpatient care.

Acknowledgments

Preliminary results from this project were presented as part of a panel presentation at the 2015 Annual Meeting of the American Medical Informatics Association. This research was supported by the National Institutes of Health, National Institute of Aging, through the Claude D Pepper Older Americans Independence Center, grant number P30AG021342 and a Career Development Award, grant number K23AG045338-05 (author SRG). This research was also sponsored by an intramural grant for Digital Health by the UCSF (no grant number, author SRG, PI). No sponsor had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript; decision to submit the manuscript for publication.

Authors' Contributions

SRG and ADA were involved in the conception of the work and interpretation of data. SRG, YM, and JR were involved in acquisition, analysis, and interpretation of data, drafting of the manuscript and revisions for important intellectual content. RJ was involved in acquisition and interpretation of data and revisions to the manuscript for important intellectual content. ADA and JDH were involved in revisions to the manuscript for important intellectual content. JDH was involved in analysis and interpretation of data and drafting of the manuscript. SRG was involved in agreement to be accountable for all aspects of the

work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. SRG, YM, JR, RJ, ADA, and JDH were involved in final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study debrief interview.
[DOCX File , 15 KB - jmir_v22i1e13337_app1.docx]

Multimedia Appendix 2
Codes and themes.
[DOCX File , 16 KB - jmir_v22i1e13337_app2.docx]

References

1. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
2. 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]
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 Mar 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
4. Otte-Trojel T, de Bont A, Rundall T, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
5. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
6. Laurance J, Henderson S, Howitt PJ, Matar M, Al Kuwari H, Edgman-Levitan S, et al. Patient engagement: four case studies that highlight the potential for improved health outcomes and reduced costs. Health Aff (Millwood) 2014 Sep;33(9):1627-1634. [doi: 10.1377/hlthaff.2014.0375] [Medline: 25201668]
7. Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
8. 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 Mar 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
9. 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 4;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
10. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 2013 Dec;28(4):1124-1139 [FREE Full text] [doi: 10.1037/a0033911] [Medline: 24364414]
11. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May;57(3):491-506 [FREE Full text] [doi: 10.1177/0018720814549238] [Medline: 25875437]
12. Wang T, Dolezel D. Usability of web-based personal health records: an analysis of consumers' perspectives. Perspect Health Inf Manag 2016;13:1f [FREE Full text] [Medline: 27134611]
13. 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]
14. Greysen SR, Khanna RR, Jacolbia R, Lee HM, Auerbach AD. Tablet computers for hospitalized patients: a pilot study to improve inpatient engagement. J Hosp Med 2014 Jun;9(6):396-399 [FREE Full text] [doi: 10.1002/jhm.2169] [Medline: 24523051]
15. Ludwin S, Greysen SR. Use of smartphones and mobile devices in hospitalized patients: untapped opportunities for inpatient engagement. J Hosp Med 2015 Jul;10(7):459-461 [FREE Full text] [doi: 10.1002/jhm.2365] [Medline: 25872902]
16. Kelly MM, Coller RJ, Hoonakker PLT. Inpatient portals for hospitalized patients and caregivers: a systematic review. J Hosp Med 2018 Jun 1;13(6):405-412 [FREE Full text] [doi: 10.12788/jhm.2894] [Medline: 29261819]
17. Collins SA, Rozenblum R, Leung WY, Morrison CR, Stade DL, McNally K, et al. Acute care patient portals: a qualitative study of stakeholder perspectives on current practices. J Am Med Inform Assoc 2017 Apr 1;24(e1):e9-17. [doi: 10.1093/jamia/ocw081] [Medline: 27357830]

18. Grossman L, Choi S, 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 1;25(4):370-379. [doi: 10.1093/jamia/ocx074] [Medline: 29040634]
19. Greysen S, Harrison J, Rareshide C, Magan Y, Seghal N, Rosenthal J, et al. A randomized controlled trial to improve engagement of hospitalized patients with their patient portals. J Am Med Inform Assoc 2018 Dec 1;25(12):1626-1633. [doi: 10.1093/jamia/ocy125] [Medline: 30346543]
20. Greysen SR, Mendoza Y, Rosenthal J, Jacolbia R, Rajkomar A, Lee H, et al. Using tablet computers to increase patient engagement with electronic personal health records: protocol for a prospective, randomized interventional study. JMIR Res Protoc 2016 Sep 6;5(3):e176 [FREE Full text] [doi: 10.2196/resprot.4672] [Medline: 27599452]
21. UCSF MyChart. UCSF Health and Affiliates Terms and Conditions for UCSF MyChart URL: https://ucsfmychart.ucsfmedicalcenter.org/ucsfmychart/default.asp?mode=stdfile&option=termsandconditions [accessed 2019-07-25]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Patton MQ. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
24. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
25. UCSF MyChart. Higher Tech and Higher Touch URL: https://www.ucsf.edu/news/2015/04/125386/higher-tech-and-higher-touch [accessed 2018-10-05]
26. Penn Medicine. The Pavilion URL: https://www.pennmedicine.org/for-patients-and-visitors/penn-medicine-locations/hospital-of-the-university-of-pennsylvania/pavilion [accessed 2018-10-05]
27. Dykes PC, Stade D, Dalal A, Clements M, Collins S, Chang F, et al. Strategies for managing mobile devices for use by hospitalized inpatients. AMIA Annu Symp Proc 2015;2015:522-531 [FREE Full text] [Medline: 26958185]
28. Al Ayubi SU, Pelletier A, Sunthara G, Gujral N, Mittal V, Bourgeois FC. A mobile app development guideline for hospital settings: maximizing the use of and minimizing the security risks of 'Bring your own devices' policies. JMIR Mhealth Uhealth 2016 May 11;4(2):e50 [FREE Full text] [doi: 10.2196/mhealth.4424] [Medline: 27169345]
29. Apple. Doctors Put Patients in Charge With Apple's Health Records Feature URL: https://www.apple.com/newsroom/2018/03/doctors-put-patients-in-charge-with-apples-health-records-feature/ [accessed 2018-10-05]
30. Dalal AK, Bates DW, Collins S. Opportunities and challenges for improving the patient experience in the acute and postacute care setting using patient portals: the patient's perspective. J Hosp Med 2017 Dec;12(12):1012-1016. [doi: 10.12788/jhm.2860] [Medline: 29073310]
31. O'Leary KJ, Lohman ME, Culver E, Killarney A, Smith G, 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]
32. Roberts S, Chaboyer W, Gonzalez R, Marshall A. Using technology to engage hospitalised patients in their care: a realist review. BMC Health Serv Res 2017 Jun 6;17(1):388 [FREE Full text] [doi: 10.1186/s12913-017-2314-0] [Medline: 28587640]
33. O'Leary KJ, Sharma RK, Killarney A, O'Hara LS, Lohman ME, Culver E, et al. Patients' and healthcare providers' perceptions of a mobile portal application for hospitalized patients. BMC Med Inform Decis Mak 2016 Sep 21;16(1):123 [FREE Full text] [doi: 10.1186/s12911-016-0363-7] [Medline: 27653854]
34. 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]
35. Dalal A, Dykes P, Collins S, Lehmann L, Ohashi K, Rozenblum R, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87. [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
36. Pell J, Mancuso M, Limon S, Oman K, Lin C. Patient access to electronic health records during hospitalization. JAMA Intern Med 2015 May;175(5):856-858. [doi: 10.1001/jamainternmed.2015.121] [Medline: 25751393]
37. Prey J, Restaino S, Vawdrey D. Providing hospital patients with access to their medical records. AMIA Annu Symp Proc 2014;2014:1884-1893 [FREE Full text] [Medline: 25954461]
38. 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]

Abbreviations

BYOD: bring-your-own-device
EHR: electronic health record
ICU: intensive care unit

RA: research assistant
RCT: randomized controlled trial
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 02.04.19; peer-reviewed by K Powell, F Bourgeois; comments to author 14.06.19; revised version received 09.08.19; accepted 19.08.19; published 14.01.20.

Please cite as:
Greysen SR, Magan Y, Rosenthal J, Jacolbia R, Auerbach AD, Harrison JD
Patient Recommendations to Improve the Implementation of and Engagement With Portals in Acute Care: Hospital-Based Qualitative Study
J Med Internet Res 2020;22(1):e13337
URL: https://www.jmir.org/2020/1/e13337
doi:10.2196/13337
PMID:31934868

Original Paper

Digitalizing Health Services by Implementing a Personal Electronic Health Record in Germany: Qualitative Analysis of Fundamental Prerequisites From the Perspective of Selected Experts

Sabrina Pohlmann[1], MSc; Aline Kunz[1], MSc; Dominik Ose[2], MPH; Eva C Winkler[3], MD, PhD; Antje Brandner[4], MSc; Regina Poss-Doering[1], MSc; Joachim Szecsenyi[1], MD, PhD; Michel Wensing[1], PhD

[1]Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany

[2]Department of Family and Preventive Medicine, University of Utah, Salt Lake City, UT, United States

[3]Department of Medical Oncology, National Center for Tumor Diseases, Research Program for Ethics and Patient-Oriented Care, University Hospital Heidelberg, Heidelberg, Germany

[4]Department of Information Technology and Medical Engineering, University Hospital Heidelberg, Heidelberg, Germany

Corresponding Author:
Sabrina Pohlmann, MSc
Department of General Practice and Health Services Research
University Hospital Heidelberg
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 6221 56 4743
Fax: 49 6221 56 1972
Email: pohlmann-sabrina@gmx.net

Abstract

Background: The implementation of a personal electronic health record (PHR) is a central objective of digitalization policies in the German health care system. Corresponding legislation was passed with the 2015 Act for Secure Digital Communication and Applications in the Health Sector (eHealth Act). However, compared with other European countries, Germany still lags behind concerning the implementation of a PHR.

Objective: In order to explore potential barriers and facilitators for the adoption of a PHR in routine health care in Germany, this paper aims to identify policies, structures, and practices of the German health care system that influence the uptake and use of a PHR.

Methods: A total of 33 semistructured interviews were conducted with a purposive sample of experts: 23 interviews with different health care professionals and 10 interviews with key actors of the German health care system who were telematics, eHealth, and information technology experts (eHealth experts). The interviews were transcribed verbatim and subjected to a content analysis.

Results: From the expert perspective, a PHR was basically considered desirable and unavoidable. At the same time, a number of challenges for implementation in Germany have been outlined. Three crucial themes emerged: (1) documentation standards: prevailing processes of the analog bureaucratic paper world, (2) interoperability: the plurality of actors and electronic systems, and (3) political structure: the lack of clear political regulations and political incentive structures.

Conclusions: With regard to the implementation of a PHR, an important precondition of a successful digitalization will be the precedent reform of the system to be digitized. Whether the recently passed Act for Faster Appointments and Better Care will be a step in the right direction remains to be seen.

(J Med Internet Res 2020;22(1):e15102) doi:10.2196/15102

KEYWORDS

personal electronic health record; implementation; qualitative analysis; digitalization; health services

Introduction

Based on many examples in Europe and beyond, the German government has been pursuing the goal to continue to expand the telematics infrastructure and the introduction of a personal electronic health record (PHR) for all insured persons for several years now [1-3]. Since 2004 and the roll out of the Statutory Health Insurance Modernization Act, the PHR has been regarded as a central feature of the envisaged telematics infrastructure [4]. Since then, starting dates for the implementation of a PHR have been postponed repeatedly.

In the literature, positive developments in European neighboring countries are highlighted as well [5]. Here the understanding of electronic health records is not always clear, both nationally and internationally. Terms and acronyms are often used synonymously and are not always clearly defined. However, meanings often differ in terms of stored data, functions, administration and access rights [6]. A study by the Muench Foundation [2] shows which countries are a nose ahead in the implementation of an electronic health record, which also includes the (patient-administered) PHR. In 2016 and in a follow-up in 2018, a study was carried out in 20 selected European countries to ascertain the status of the introduction of a (personal) electronic health record [2]. On the "European Scorecard," Germany was shown ranking in the lower midfield. In 2018, Germany almost slipped into the bottom third of the countries surveyed. On the other hand, the Scandinavian countries are the top scorers [2]. While in other countries PHRs have obviously already been established as a central component of national eHealth strategies, in Germany the implementation and application seem to be confronted with certain hurdles. The PHR is regarded as the "supreme discipline" [7] of the digitalization of the German health care system. A recent study analyzed conditions for its success, pointing to central barriers for digitalization in health care [8]. Hesitance with the implementation of a PHR in Germany has been attributed to political framework conditions as a missing or incomplete prerequisite [2,4]. At the same time, the need to relate the implementation more strongly to practical everyday needs and questions of the respective users is propagated [3].

The aim of this study was to explore potential barriers and facilitators for the adoption of a PHR in routine health care in Germany. The PHR here was specified as a Web-based personal electronic health record prototype (Persönliche, einrichtungsübergreifende, elektronische Patientenakte [PEPA], also called a personal electronic health record [PEHR]), where the patient is the owner of their health record [9,10]. This paper focuses on reasons for the failure to implement a PHR in Germany and ventures some ideas to overcome identified obstacles. Therefore, it explores relevant policies, structures, and practices of the German health care system that influence the uptake and use of a PHR.

Methods

This study was based on semistructured and semistandardized interviews with potential users and eHealth experts. Reporting of this study follows the recommendations of the consolidated criteria for reporting qualitative research checklist (COREQ) [11].

Study Design

The study is part of the Information Technology for Patient-Oriented Health Care in the Rhine-Neckar Region project, which was funded by the German Federal Ministry of Education and Research (2012-2017) and focused on the structural prerequisites for integrated and cross-sector care of chronically ill people and the scale-up of the prerequisites. A central component of this project was the development of a PHR for use in cancer patients [9,12,13]. Two key features of the PHR are that it is Web-based and patient-administered. Patient-administered means that the patient decides who has access to which parts of the recorded information [12]. Central to the development and application of the PHR is a strong focus on the needs and preferences of end users [9,14]. Based on results from implementation research [15,16], the main objective of the study was to realize a strong user centering and thus include the specific needs and requirements of all PHR user groups in an iterative process. In addition, both the action and work contexts of the user groups as well as relevant structural prerequisites were included in the analysis.

This article focuses on the findings of a subproject in which a usability and feasibility study was conducted. Here the prototype of the PHR was tested and refined with various eHealth experts, health care professionals, and cancer patients involved. In addition to the technical usability, relevant policies, structures, and practices for a successful implementation of a PHR were evaluated [9].

Ethical approval was given by the Ethics Committee of the University of Heidelberg (S-462-2015). All participants gave their written informed consent. The participants' anonymity and confidentiality were ensured throughout the study.

Study Sample

E-Health Experts

Participants had to be central stakeholders of the German health system or experts in the fields of information technology, telematics, and eHealth. These included organizations like political institutions, federal associations, chambers of commerce, and hospital management. The sampling of participating experts was based on (1) their thematic interest, (2) position or reputation of the specific expert, and (3) potential impact to foster political decisions for a broader PHR implementation. A total of 14 eHealth experts from 14 different institutions were contacted; 3 experts decided not to participate. One expert could not participate in the study for health reasons on short notice. In 4 cases, the contacted persons could name another person in their institution who instead took part in the study. The eHealth experts were recruited by the Department of General Practice and Health Services Research (University Hospital Heidelberg) via postal mail.

Health Care Professionals

Based on an intersectoral care setting for cancer patients with colorectal carcinoma developed in this project context, central health care professional groups could be identified and recruited.

The study included hospital physicians, nutritionists, and social service workers of the National Center for Tumor Diseases (NCT) in Heidelberg as well as general practitioners (GP) and nonphysician health professionals who work together in the care context of patients with colorectal cancer [9]. All interviewees were recruited using a purposive sampling method. HCPs from the Rhine-Neckar area were contacted according to the research interests. They were identified and selected as stakeholders of PHRs from our research practices and our cooperation partner (NCT). HCPs were recruited by the Department of General Practice and Health Services Research (University Hospital Heidelberg). They were approached face to face or by email.

Data Collection

The semistandardized and semistructured interviews were conducted in two survey periods. Interviews with the HCPs took place from September 2015 to March 2016. Interviews with the eHealth experts take place between June and October 2016. Interview duration was 30 to 60 minutes.

Prior to the interviews, the HCPs participated in a usability test with the developed PEPA prototype that lasted an average of 35 minutes. Here the HCPs could gather tangible experiences in handling PHRs and formulate idea on positive as well as negative aspects, both in general and specific to their own work context. The focus here was on technical criteria directly related to the manageability of the PHR. The following interviews were expected to generate statements about prerequisites and hurdles beyond pure usability. The usability testing facilitated alignment of barriers and requirements with PHR. The interviews were based on a semistandardized and semistructured pilot-tested interview guide that focused on the main areas of prerequisites and challenges. Themes and questions in this interview guide were based on theoretical considerations and findings from a literature review [17]. The interviews had a duration of 30 to 45 minutes. All but one was conducted in the respective work environment.

Since the experts did not participate in a usability test, a short description of the PHR and access to the PHR internet portals (professional portal for HCPs; patient portal for patients) were sent to the experts prior to the interview. Thus, an overview of the functionality and structure of the PEPA as well as the objectives of the PEPA concept were provided to all interview partners in advance. If requested, a question catalog was sent to the experts beforehand.

The interviews were generally conducted face to face by the first author (SP) at the interviewees' workplace. Two interviews were conducted by telephone. The interviews lasted 45 to 60 minutes. All interviews were audiotaped and transcribed verbatim but not returned to the participants. The interviews lasted until the saturation of theoretical arguments was reached.

Data Analysis

The performed qualitative content analysis explored pseudonymized and purged data taken from the original texts [18-20]. A preliminary category system was applied based on theoretical considerations. The category system was adapted during the process of analysis if the data showed additional and new information that required a new category. Therefore, analysis included inductive development as well as deductive application of categories.

In the first step, two interview transcripts from each participant group (HCPs and eHealth experts) were chosen and independently reviewed by the first author (SP), a coauthor (AK), and two other scientific colleagues who were not involved in the project but experienced in qualitative evaluation methods. In the second step, further central topics could be developed, summarized, and labeled based on the temporary category system. The resulting defined codes were differentiated into main and subcategories. Relevant original statements in the transcripts were assigned to these categories. Referencing the developed category system, attributions were repeatedly adapted and condensed in a 3-step discussion procedure with a coauthor (MW). Qualitative content analysis of the data was performed using the software MAXQDA Analytics Pro 18 (VERBI GmbH).

Results

Sample Description and General Picture

The study included a total of 10 eHealth experts and 23 HCPs. The eHealth experts were key players in the German health care system working in information technology and telematics at the levels of federal politics, self-administration (federal/state—in Germany, the state sets the legal framework and assignments and the insured people, contributors, and HCPs organize themselves into associations that assume the medical care of the German population on their own responsibility), other associations at the federal level, and hospital management. The HCPs comprised three groups: general practitioners (6), hospital physicians (5), and other health care professions (nonphysician health professional, nutritionists, social service workers; 12).

To gain a general picture of the mood regarding PEPA and its fundamental application in Germany, three pivotal questions were asked in the expert interviews. All the eHealth experts found the application of a PHR generally desirable. When asked whether the prerequisites for an implementation of PHR had already been met, 4 interviewees answered with "rather no," 3 with "partly," and another 3 with "rather yes." This indicates that most of the eHealth experts did not yet or only partially see the most important prerequisite for a successful PHR implementation. When asked whether a PHR could be realistically assessed as a regulatory instrument in 5 to 10 years, almost all respondents answered yes. HCPs also showed a very high acceptance of PEPA. All of the hospital physicians and most of GPs stated that they would use a PHR on a regular basis.

Central prerequisites and associated challenges could be categorized into 3 main topics: (1) documentation standards: prevailing processes of the analog bureaucratic paper world, (2) interoperability: the plurality of actors and electronic systems, and (3) political structure: the lack of clear regulations and incentive structures.

Documentation Standards: Prevailing Processes of the Analog Bureaucratic Paper World

For most eHealth experts, existing mindsets are still not well matched with digitalization processes. Established procedures of the analog paper world would be prevailing, and electronic duplicates of bureaucratic forms of health care services were leading into the wrong direction. For them it went without saying that the electronic world cannot and should not simply duplicate the paper world of the bureaucracy. If just mirroring bureaucratic forms, the implementation is doomed to failure.

> *...this can't work like that. We have a highly structured system that is often perceived as much too bureaucratic, based on the paper world ...all the template forms that exist in [statutory health insurance]–accredited medical care, they all come from the template form commission. They have been decided on, and every field is being discussed...it is impossible and I do not think it makes any sense.... So, you don't do justice to the patients or their data....* [E8]

> *...currently there is a template, don't know what it is called, 37b or something like that, in triple or quadruple copy, green, blue, pink, or something else... it's not about mapping this process by a digital document in an electronic health record, but it's about digitizing the process.* [E17]

For the eHealth experts, this approach to digitalization was considered wrong and one of the biggest mistakes in the introduction of a PHR. For them, digitalization requires a different logic of how to do things, how to establish standard operation procedures. If analog bureaucratic rationality takes the lead, the project is jeopardized in this perspective. So, to make a PHR a success, existing practices in health care have first to be rationalized, simplified, and redesigned. The right approach, according to the experts, would then be to set up digital structures. A template form with four carbon copies simply should not be duplicated electronically. According to the message of the experts, anyone who stays in the paper world mentally will not be able to realize the advantages of the digital world.

> *The eHealth Act states that the self-governing body has been commissioned to study the Social Security Statutes (SGB V) and based on its name which of the many forms and so on and so forth could be digitally reproduced. This in a digital age! Basically, it would have to be the other way around, meaning the order would have to be, everything is done digitally, unless there are substantial reasons impeding going digital. Well, that's the first reason it doesn't work...we do not think digitally. We think within our analog processes and believe that with a health record, we can electronify the analog processes. That is not the purpose of a health record.* [E17]

This problem in dealing with digitalization also became evident in a more concrete form at the level of health organizations. HCPs also addressed the multitude of different forms that cannot just be mirrored in the electronic world. It was mentioned by HCPs that the specialized, different documentation standards, especially for clinicians, specialists, and GPs, could constitute a serious barrier to general standards of electronic documentation.

> *So, my GP still has her paper-based records and if I imagine she has to (change) that, oh God, that would be difficult.* [HP02]

> *They also have different documentation, different templates [the clinics].* [GP01]

Also, documentation is often divided into internal documentation not intended for the patient and external formal documentation. To bring this into a comprehensible and accessible version seemed impossible to many.

> *...that the patient can read this, not every report is suitable for the patient.* [GP02]

> *Yes. You have some things you just don't want the patient to read now, because it's really something internal, and that will be difficult, of course.* [HP02]

In principle, almost everyone was concerned that they would have to deal with a double structure and thus a double workload. Especially since the systems used are tailored to these technical differences and support the differently standardized documentation.

> *...so, such a double structure that would result, for example, inevitably, that makes double work, so that throws the baby out with the bath water then.* [GP03]

In this context, the problem of proprietary systems, which are usually incompatible, was often mentioned as well. In almost all cases, physicians agreed to an electronic health record, but when more precise inquiries were made, very often restrictions were mentioned regarding the misleading approach of digitalization. Again, the paper world was not to be duplicated, but a reform of the system deemed necessary to digitize it.

Interoperability: Plurality of Actors and Electronic Systems

Most eHealth experts agreed that it is a prerequisite for successful implementation of a PHR to reach standards of interoperability. But because of the plurality of electronic systems and actors, including the self-governing bodies, important steps in that direction have not been undertaken so far. Instead of addressing the challenge of interoperability in an accurate manner, isolated solutions were said to be prevailing in Germany. That was mentioned by interviewees when they addressed the status quo of electronic health record applications. It was stated that there were many isolated solutions in Germany including regional projects. The interviewees regretted that it was still impossible to draw these projects onto a supraregional level. The interviewees also criticized the fact that physicians do not have access to the health telematics system. In this context, most of the interviewees emphasized that interoperability must be a first step toward introducing health telematics.

> *...but if the doctors do not have access to health telematics, then one cannot introduce a health record. The health records are not interoperable, neither*

> *technically nor semantically interoperable, and we have a huge problem with that.* [E10]

The majority of the eHealth experts addressed interoperability as a relevant aspect for the introduction and implementation of a PHR. Basically, almost all interview partners said that there were a multitude of electronic file systems, but in the end interoperability did not exist. This pointed to proprietary business models seen as representative of a fundamental problem for the implementation of a PHR.

> *...and this also has to do with these proprietary solutions, of course, because the more proprietary systems you have to begin with, the more problems you have as a user. And it doesn't go back to the fact that the IT technically can't be solved. But it goes back to the fact that old business models are still very strong and very present there, and in the end prevent a good workflow from developing there. And that is simply something, in my opinion, one cannot somehow preach to the physicians from morning to evening that these IT systems are the future, if they experience in their practice every day that this IT causes problems and doesn't solve any at all.* [E16]

As the first developments of digitalization in health services were focused on billing purposes and the digital storage of billing data and processes in general, care and digital processes relevant to care were not considered.

> *So, one simply does not have the focus on structured filing of medical documents but one has the focus on uniform evaluation standard numbers (EBM-Ziffern)...and if the systems are knitted in this way, then they cannot quickly communicate interoperably with a third-party system....* [E8]

Most of the HCP interviewees agreed that there are severe problems with integrating different electronic systems. It is said to be a mix of paper and pencil procedures and electronic file systems. Before implementation of a PHR could take place, the uncontrolled growth of different standard operation procedures in the file systems would have to be terminated.

> *...so I think that...the individual clinic systems or practice systems should be better coordinated with each other.... And, yes, I don't know how far away the individual providers are from participating, everyone prefers to market their product....* [HP10]

Political Structure: Lack of Clear Regulations and Incentive Structures

Political Regulations

Many of the eHealth experts and HCPs were arguing in favor of a strong central solution for the problems. Only the legislature would be responsible for providing interoperability. In their mindsets, it was necessary to create legislation and give a clear mandate on who would realize implementation and take responsibility in this context.

> *...from my point of view, the decisive actor is the Federal Ministry of Health.* [E12]

> *I believe that interoperability can only come directly from the legislature. The self-governing body simply has no mandate for this, not derived from the law. In my opinion, the legislature has very hesitantly included interoperability in the eHealth Act. However, this is limited to an archiving interface for patient data-bearing systems. This is a start but is of no use with regard to the health record. The legislature would have to prescribe this directly and then give the order...so to speak provided with this mandate, then specifications gladly can be made together with the industry also....* [E8]

In addition to interoperability as a central prerequisite for the digitalization of care in general and implementation of a PHR in particular, the clear political will, corresponding laws, procedures, and incentive structures were emphasized. Top-down solutions were generally preferred and seen as harmful only by a minority of eHealth experts. A large number of respondents see that the demand for PHR should originate from patients. It is then the task of politicians, and almost all experts agree on this, to moderate this process and create the necessary framework conditions for it.

> *I need data protection regulations, I need all the rights and obligations of physicians and service providers, I need a reasonable remuneration system for this, and, in particular, I need the corresponding processes within the health care system behind it: I do that now, and I say:...I'll push it through and I'll put it into practice. So, anyone who wants to join is cordially invited to do so, anyone who doesn't want to join will be left out... Constructive suggestions are welcome; destructive suggestions lead to exclusion. That's it. And that [the source of these processes] can only be the legislature....* [E17]

> *...the patients have to be involved decisively; ...the patient has to play a decisive role.* [E16]

Almost all interviewees considered the political will on a societal level imperative for the acceptance of a PHR on a broad social basis. Here, the focus was on greater transparency with regard to the discussed contents and decisions. At the same time, citizens should be directly involved in the decision-making processes and the content discussed.

> *...the question [is], how transparent are the decision-making procedures...that this takes place in a quiet chamber, that the discussion processes are not made public, that is quite unpleasant.... So, I believe that it is much more important to get a higher transparency of the negotiations....* [E16]

According to the interviewees, the central and first task of the state is to ensure that the patient's fundamental right to determine disclosure and use of their personal data is respected and legally strengthened. More than half of the eHealth experts considered this aspect to be a central prerequisite for the design and implementation of a PHR.

> *Well, those are the big challenges, but I think this right to determine disclosure and use of personal data, that's right at the top....* [E6]

> *Basically, I would say that in Germany you sometimes protect the citizen too much from themselves. In this case, I consider the right of informational self-determination to be of more valuable and would in any case plead for it....* [E10]

At the same time, this was also central to the question of who owns the data. It was propagated that the ownership rights of the data—personal information about and from the patients—must be clarified and defined.

> *Well, first of all the prerequisite [for] the application of an electronic personal health record is, in my opinion...the patient's right to data sovereignty.* [E17]

A similar picture emerges with the HCPs. Many of them also argue that legislatures need to create clear guidelines and responsibilities, especially when it comes to the question of the ownership rights to the data.

> *Well, the Federal Ministry of Health [certainly] would have to put it on the agenda: "We want that." Otherwise it will be difficult.* [GP08]

> *One thing I have learned in my life: without legislation nothing works at all. Thus, the legislature is challenged. ...[Also] in the context of all these discussions about patient autonomy.... The framework conditions must be created by the legislature. That the patient data...is usually not only in the physician's practice, but with the patient....* [GP01]

For most experts, the introduction of a PHR was based on a clear, binding political decision with the allocation and definition of clear responsibilities. Almost all of them pointed to the problem of the normal design and procedures of the health care system in Germany. Specifically, the negotiation habits and procedures of self-administration were addressed.

Incentive Structures

With regard to the strongly pronounced sectoral separation in Germany, existing incentive systems within the current system and, above all, false incentives for the introduction of electronic health records were highlighted as relevant obstacles.

> *A clear specification of the Minister of Health...: "Until then it must come, otherwise there are financial deductions." Or one could work with a bonus-malus [principle]... Probably this is the only language that hospitals...[and also the outpatient sector understand]. ...One declares: "Until then it must be implemented," period.* [E01]

The difficulty of placing this issue in the hands of self-government was pointed out, since in many respects the usual negotiation routines were considered to make the actual process more difficult. One interviewee described this as a "playground" for the "usual games and trench wars."

> *The cake must be distributed...these digitalization processes are such a vehicle. So, it's a new playground that's being used for that.* [E6]

In order to keep PHR systems running, carrot and stick policies on the provider's side were contemplated as well. Interviewees mentioned the various remuneration systems and incentive structures that in their opinion would change the mindset of those HCPs who do not support the implementation of a PHR.

> *I have come to the point where I believe that certain processes, certain digitalization processes cannot be left to the regular structures of our health system. ...This is such a profound infrastructural change that I believe it must be pushed through relatively hard by the state with a carrot and stick system.* [E6]

One interviewee described the regulation of performance reimbursement as "the manifestation of isolation." Due to the different performance reimbursements, conduct of the individual HCP was seen as being structured in a way that imposes a severe obstacle for the implementation of the PHR. Different fees thus would determine to what extent and in what way communication between HCPs is desired and in what direction it would take place.

> *Well, when the data flows, there's always the danger...that the money flows afterwards. That is what the sectors are striving for. ...And—who has the data, has the power.* [E6]

In this context, the sender-receiver problem was addressed, in which the sender is usually the disadvantaged person, while the receiver attains the greater benefit from digitalization, which must be compensated for. This means that one part of the medical profession feels more disadvantaged than another.

> *We often have an imbalance in benefits, so that we have to somehow create incentives or a balance in order to start the whole thing in the first place. ...Especially [among] GPs and specialists, because some always see themselves as transmitters and others as receivers. Then you get into a situation where the physicians, who were originally able to work well together, first see themselves as opposing actors. Because you simply haven't created the setting that gives them the feeling, to take this digitalization step equally.* [E8]

Not only will a clear position of the politicians on this topic be expected, but also clear guidelines up to the presentation of a model process (eg, for hospitals).

> *So, one [the legislature] should prestructure the whole process unanimously also times as a pattern [process].* [E1]

Discussion

Principal Findings

The study aimed to explore potential barriers and facilitators for the adoption of a PHR in routine health care in Germany. This paper focuses on reasons for the failure to implement a PHR in Germany and ventures some ideas to overcome identified obstacles. Therefore, it explores relevant policies, structures, and practices of the German health care system that influence the uptake and use of a PHR.

The results showed that organizational, economic, and political interrelations become relevant. Three themes emerged to be central to the implementation of a PHR in Germany: (1)

documentation standards: prevailing processes of the analog bureaucratic paper world, (2) interoperability: the plurality of actors and electronic systems, and (3) political structure: the lack of clear regulations and incentive structures (Textbox 1).

Textbox 1. Principal findings: categories and themes.

Documentation standards:

- Highly segmented health care
- Mandate of the analog paper world
- Highly bureaucratic
- Strongly differentiated
- Various standards and terminologies

Interoperability:

- Federal structures
- Highly segmented health care
- Plurality of electronic systems
- Old business models
- Plurality of actors
- Semantic and technical interoperability

Political structure:

- Clear political will
- Incentive structure
- Rights and regulations
- Self-administration
- Different interests
- Strong advocacy groups
- Clear responsibilities
- Transparency

Documentation Standards: Prevailing Processes of the Analog Bureaucratic Paper World

Starting from a highly segmented care landscape, the various standards in medical documentation represent a central challenge for the implementation of a cross-sectoral application of a PHR. The differentiated use of documentation can therefore often be attributed to the differentiation between inpatient and outpatient care, as well as to the various specializations. The result is a multitude of distinguished forms and heterogeneous terminologies. In literature, this is often discussed in the context of semantic interoperability, which is considered a prerequisite to be able to guarantee the targeted exchange of patient-related documentation at all [21,22]. The highly bureaucratic procedures in dealing with documentation make the transformation of the analog paper world into the electronic form of a PHR considerably more difficult. If patients are included in the exchange of information, these challenges were specified in terms of health literacy [23].

It will be important to take these prerequisites into account when designing the digitalization of the analog paper world. The pure duplication of the most varied documentation standards and approaches from the paper world continues to lead to challenges for the creation of semantic interoperability [21,22]. Rather, it becomes necessary to adopt digital processes. As literature in the context of digital medicine has shown, a common technical language as well as concrete guidelines for the exchange formats of this data will be inevitable for a successful adoption of digital technologies [24]. A reform and reorganization of previous standards and processes as well as a different understanding of digitalization will be unavoidable. Different administrative practice and clinic systems only reproduce the variety and diversity of existing standards. The challenges already mentioned for the documentation behavior of physicians with regard to the use of a PHR [25-27] are partially confirmed by our results (eg, in the case of documentation quality, double structure/double documentation standards, misinterpretations). The current market situation for electronic health records reflects this diversity of existing standards and the difficulty of linking them.

Interoperability: Plurality of Actors and Electronic Systems

In the context of a segmented and federal health care landscape with a variety of different documentation systems and a

multitude of providers, interoperability is a major challenge. Interoperability can be understood as the ability of independent, heterogeneous systems to work together as seamlessly as possible in order to exchange information in an efficient and usable manner or to make it available to the user [21]. Statement papers in the political discussion generally propagate the development of interoperability as a central condition for the application of a PHR: "We want to create new approval paths for digital applications, establish interoperability, and strengthen digital security in the health care system" [1]. Interoperability stands for overcoming the often portrayed island solutions, which have so far been predominant in Germany [4].

The existence of a multitude of electronic systems and proprietary business models makes it difficult to implement interoperable systems. Uniform technical standards and interfaces for the exchange of data cannot yet be guaranteed in the way necessary for the use of a PHR. In order to enable digital structures and practices, as it is necessary in the case of PHR, the literature speaks, among other things, of creating uniform industry standards to enable interoperable systems [2,8].

However, as the findings of this study show, it seems not to be enough to recognize and solve interoperability as a purely technical challenge. Semantic and technical interoperability must therefore go hand in hand [21], which entails a far-reaching reorganization of the bureaucratic paper world and the associated restructuring of structures and practices in health care. Clarifying the market situation by creating a telematics infrastructure will be as important as standardizing and simplifying documentation methods.

Political Structure: Lack of Clear Political Regulations and Political Incentive Structures

The typical self-administration structure of the German health care system enables strong stakeholder participation in decision-making processes [28]. Similarly, as in the case of a PHR, there is a risk that negotiation routines and diverging interests will hamper a clear line in the implementation and design of a PHR. As propagated from the experts, a greater legislative responsibility could speed up the implementation process and ensure a consistent and clear approach. Also, in literature, a uniform policy strategy with clear responsibilities as well as a greater transparency for citizens and their involvement is needed for a sustainable digitalization of health care and implementation of a PHR [8]. From an expert's point of view, it is necessary to develop an overall strategy based on the eHealth Act, particularly in a complex self-administration structure like in Germany. This also includes the formulation and allocation of clear responsibilities with clearly framed tasks and objectives [2,8]. A relationship management [8] that bundles and moderates the various interests, goals, and opportunities is equally proposed. Some literature provides incentives for the use of a PHR for the respective user groups [8]. As a rule, the focus here is on monetary incentives, which should be available for doctors, as well as sanctions, if the expectations of use cannot be fulfilled [8]. The experts mentioned that necessary negotiations between relevant stakeholders of the health care system should include the persisting differences into existing incentives. Compensatory incentive and sanction structures should be created to enable potential user groups to benefit from PHR on a lending basis.

Confirming findings of international studies in which data protection raised only moderate concerns among potential user groups [6,24-26], data protection does not play a significant role for the interviewed experts in this study either. Even though data protection was described as too outdated and not dynamic enough by some voices, especially with regard to international standards, due to the German special way it does not represent a major challenge for the implementation of the PHR. In fact, experts believed that data protection is often misused as an argument in connection with digitalization in the health care sector in order to prevent things from happening or to avoid having to discuss relevant and more difficult issues.

Strengths and Limitations

Besides the early and consistent involvement of end users and their action and working contexts in the development of the PHR, it was another strength of this study to include the general policies, structures, and practices of the German health care system in the evaluation. Thus, central conditions of the German health care system can be examined at different levels, and their relevance for the implementation of PHR can be contemplated.

As explained above, the understanding of a PHR varies both internationally and nationally. Although before the interviews with the experts there were efforts to present the PHR concept defined in the project context in written form, the arguments of the experts can also refer to other concepts of a PHR. Even though fundamental questions on the PHR concept were asked or answered before the interviews, it cannot be conclusively ensured that the answers relate exclusively to this concept.

The fact that all HCPs included in the study work in the field of colorectal cancer may restrict the generalizability and transferability of the results to other (medical) settings.

Conclusions

For the eHealth experts in Germany, a PHR is basically desirable and unavoidable. At the same time, a number of challenges for implementation in Germany have been outlined which can be taken into future focus.

Whether the recently adopted Act for Faster Appointments and Better Care (appointment and care law, TSVG), which came into force on May 11, 2019, is a step in the right direction remains to be seen. It obliges all health insurance companies to offer electronic patient records for their insured by 2021 at the latest. In addition, the legislature assumes primary responsibility related to the creation of defined interfaces to enable interoperability. It will depend, among other things, on how the market situation clarifies itself and how patient autonomy can be strengthened under the given conditions.

In principle, it will be important to consider existing structures of the medical care landscape and the effects of digitalization processes on these structures when introducing the PHR. Especially with regard to the implementation of a PHR, one important precondition of a successful digitalization will be the precedent reform of the system to be digitized.

Acknowledgments

We would like to thank all interview partners who took the time to contribute to this topic. We would also like to thank our project partners Department of Information Technology and Medical Engineering, University of Heidelberg, and Heidelberger Selbsthilfebüro, an umbrella organization for patient support groups in Heidelberg as well as our cooperation partner, NCT, University of Heidelberg, for their collaboration. This study was funded by the German Federal Ministry of Education and Research (funding code 01KQ1003B).

Conflicts of Interest

None declared.

References

1. Presse- und Informationsdienst der Bundesregierung. Koalitionsvertrag vom 14. 2018 Mar. URL: https://www.bundesregierung.de/resource/blob/975224/847984/5b8bc23590d4cb2892b31c987ad672b7/2018-03-14-koalitionsvertrag-data.pdf?download=1 [accessed 2019-03-14]
2. Projektbericht European Scorecard zum Stand der Implementierung der elektronischen Patientenakte auf nationaler Ebene: Erstes Update. 2018. URL: https://www.stiftung-muench.org/wp-content/uploads/2018/09/Scorecard-final.pdf [accessed 2018-11-27]
3. Ex P, Amelung VE. Auf und Ab: Der politische Wille zur Stärkung der Digitalisierung im Gesundheitswesen. GuS 2018;72(2):26-30. [doi: 10.5771/1611-5821-2018-2-26]
4. Krüger-Brand H, Osterloh F. Deutsches Ärzteblatt. 2017. Elektronische Patientenakte: Viele Modelle—noch keine Strategie. URL: https://www.aerzteblatt.de/pdf.asp?id=194137 [accessed 2018-11-27]
5. Wallenfels M. Dänemark bleibt E-Akten-Europameister: Dänemark, Finnland und Schweden machen vor, wie E-Patientenakten zu einer Erfolgsgeschichte werden können. Deutschland rangiert im 20-Länder-Vergleich im hinteren Mittelfeld. Eine Besserung ist (noch) nicht in Sicht. 2018. URL: https://www.aerztezeitung.de/praxis_wirtschaft/e-health/article/973029/laendervergleich-daenemark-bleibt-e-akten-europameister.html [accessed 2019-03-13]
6. Bertram N, Püschner F, Gonçalves A, Binder S, Amelung V. Einführung einer elektronischen Patientenakte in Deutschland vor dem Hintergrund der internationalen Erfahrungen. In: Klauber J, Geraedts M, Friedrich J, Wasem J, editors. Krankenhaus-Report 2019: Das Digitale Krankenhaus. Berlin: Springer; 2019:3-16.
7. Krüger-Brand H. Die Patientenakte als "Königsdisziplin." Heft 18. 2017 May 5. URL: https://www.aerzteblatt.de/pdf.asp?id=188480 [accessed 2019-11-18]
8. Nohl-Deryk P, Brinkmann JK, Gerlach FM, Schreyögg J, Achelrod D. [Barriers to digitalisation of healthcare in Germany: a survey of experts]. Gesundheitswesen 2018 Nov;80(11):939-945. [doi: 10.1055/s-0043-121010] [Medline: 29301149]
9. Ose D, Kunz A, Pohlmann S, Hofmann H, Qreini M, Krisam J, et al. A personal electronic health record: study protocol of a feasibility study on implementation in a real-world health care setting. JMIR Res Protoc 2017 Mar 02;6(3):e33 [FREE Full text] [doi: 10.2196/resprot.6314] [Medline: 28254735]
10. Brandner A, Schreiweis B, Aguduri LS, Bronsch T, Kunz A, Pensold P, et al. The patient portal of the personal cross-enterprise electronic health record (PEHR) in the Rhine-Neckar Region. Stud Health Technol Inform 2016;228:157-161. [Medline: 27577362]
11. 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]
12. Baudendistel I, Winkler E, Kamradt M, Brophy S, Längst G, Eckrich F, et al. The patients' active role in managing a personal electronic health record: a qualitative analysis. Support Care Cancer 2015 Sep;23(9):2613-2621. [doi: 10.1007/s00520-015-2620-1] [Medline: 25652149]
13. Baudendistel I, Winkler E, Kamradt M, Längst G, Eckrich F, Heinze O, et al. Personal electronic health records: understanding user requirements and needs in chronic cancer care. J Med Internet Res 2015;17(5):e121 [FREE Full text] [doi: 10.2196/jmir.3884] [Medline: 25998006]
14. Kunz A, Pohlmann S, Heinze O, Brandner A, Reiß C, Kamradt M, et al. Strengthening interprofessional requirements engineering through action sheets: a pilot study. JMIR Hum Factors 2016 Oct 18;3(2):e25 [FREE Full text] [doi: 10.2196/humanfactors.5364] [Medline: 27756716]
15. Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, et al. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci 2013 Mar 23;8:35 [FREE Full text] [doi: 10.1186/1748-5908-8-35] [Medline: 23522377]
16. 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]

17. Ose D, Baudendistel I, Pohlmann S, Winkler EC, Kunz A, Szecsenyi J. Personal health records on the Internet. A narrative review of attitudes, expectations, utilization and effects on health outcomes. Z Evid Fortbild Qual Gesundhwes 2017 May;122:9-21. [doi: 10.1016/j.zefq.2017.04.006] [Medline: 28499764]
18. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. Weinheim: Beltz Verlagsgruppe; 2010.
19. Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. 3. überarbeitete Auflage. Weinheim: Beltz Juventa; 2016:978-983.
20. Gläser J, Laudel G. Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. 4. Auflage. Wiesbaden: VS Verlag; 2010.
21. Pedersen S, Hasselbring W. Interoperabilität für Informationssysteme im Gesundheitswesen auf Basis medizinischer Standards. Informatik—Forschung und Entwicklung 2004 Apr 1;18(3-4):174-188. [doi: 10.1007/s00450-004-0146-8]
22. Heinze O, Bergh B. Persönliche einrichtungsübergreifende Gesundheits- und Patientenakten (PEPA) als zentrale Infrastrukturkomponente einer patientenzentrierten Gesundheitsversorgung. In: Müller-Mielitz S, Lux T, editors. E-Health-Ökonomie. Wiesbaden: Springer Fachmedien; 2017.
23. Showell C. Barriers to the use of personal health records by patients: a structured review. PeerJ 2017;5:e3268 [FREE Full text] [doi: 10.7717/peerj.3268] [Medline: 28462058]
24. Thun S. Digitalisierte Medizin. Informatik Spektrum 2014 Dec 18;38(1):22-27. [doi: 10.1007/s00287-014-0859-4]
25. Walker J, Leveille SG, Ngo L, Vodicka E, Darer JD, Dhanireddy S, et al. Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. Ann Intern Med 2011 Dec 20;155(12):811-819 [FREE Full text] [doi: 10.7326/0003-4819-155-12-201112200-00003] [Medline: 22184688]
26. Yau GL, Williams AS, Brown JB. Family physicians' perspectives on personal health records: qualitative study. Can Fam Physician 2011 May;57(5):e178-e184 [FREE Full text] [Medline: 21642732]
27. Wynia MK, Torres GW, Lemieux J. Many physicians are willing to use patients' electronic personal health records, but doctors differ by location, gender, and practice. Health Aff (Millwood) 2011 Feb;30(2):266-273 [FREE Full text] [doi: 10.1377/hlthaff.2010.0342] [Medline: 21289348]
28. Böhm K, Schmid A, Götze R, Landwehr C, Rothgang H. Five types of OECD healthcare systems: empirical results of a deductive classification. Health Policy 2013 Dec;113(3):258-269. [doi: 10.1016/j.healthpol.2013.09.003] [Medline: 24095274]

Abbreviations

COREQ: consolidated criteria for reporting qualitative research checklist
GP: general practitioner
HCP: health care professional
NCT: National Center for Tumor Diseases
PEHR: personal electronic health record
PEPA: Persönliche, einrichtungsübergreifende, elektronische Patientenakte
PHR: personal electronic health record

Edited by G Eysenbach; submitted 19.06.19; peer-reviewed by D John, P Nohl-Deryk, C Showell, S Manzanero; comments to author 01.09.19; revised version received 01.10.19; accepted 22.10.19; published 29.01.20.

Please cite as:
Pohlmann S, Kunz A, Ose D, Winkler EC, Brandner A, Poss-Doering R, Szecsenyi J, Wensing M
Digitalizing Health Services by Implementing a Personal Electronic Health Record in Germany: Qualitative Analysis of Fundamental Prerequisites From the Perspective of Selected Experts
J Med Internet Res 2020;22(1):e15102
URL: https://www.jmir.org/2020/1/e15102
doi:10.2196/15102
PMID:32012060

Review

A Web-Based Dementia Education Program and its Application to an Australian Web-Based Dementia Care Competency and Training Network: Integrative Systematic Review

Anne Moehead[1*], MA; Kathryn DeSouza[1*], BA; Karen Walsh[1*], MA; Sabrina W Pit[2,3*], MSc, PhD

[1]Northern New South Wales Local Health District, New South Wales Ministry of Health, Lismore, Australia

[2]Western Sydney University, Sydney, Australia

[3]University Centre for Rural Health, University of Sydney, Sydney, Australia

[*]all authors contributed equally

Corresponding Author:
Anne Moehead, MA
Northern New South Wales Local Health District
New South Wales Ministry of Health
Uralba St
Lismore, 2480
Australia
Phone: 61 429449591
Fax: 61 266202180
Email: gmo11800@dodo.com.au

Abstract

Background: Dementia education that meets quality and safety standards is paramount to ensure a highly skilled dementia care workforce. Web-based education provides a flexible and cost-effective medium. To be successful, Web-based education must contain features that promote learning and support knowledge translation into practice. The Dementia Care Competency and Training Network (DCC&TN) has developed an innovative Web-based program that promotes improvement of the attitudes, knowledge, skills, behavior, and practice of clinicians, regardless of their work setting, in order to improve the quality of life for people living with dementia.

Objective: This review aims to (1) determine the key features that are associated with an effective and functional Web-based education program—an effective and functional Web-based program is defined as one that measures results, is accessible, is user friendly, and translates into clinical practice—and (2) determine how these features correlate with the DCC&TN.

Methods: Six electronic databases—Medline, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL), AusHealth, Nursing@Ovid, and Google Scholar—were searched for articles published between 2009 and 2018 using the following keywords: Education, Distance, Continuing, Learning, Online, Web-Based, Internet, Dementia, Program Evaluation, Validation Studies, Outcome and Process Assessment Healthcare, Nursing, Assisted Instruction, and Facilitated. The Critical Appraisal Skills Programme (CASP) and Kirkpatrick's model for the evaluation of training were used to ensure quality and rigor of the analysis.

Results: A total of 46 studies met the inclusion criteria. In total, 14 key features were associated with an effective Web-based learning environment, which enabled the environment to be as follows: self-directed, individualized, interactive, multimodal, flexible, accessible, consistent, cost-effective, measurable with respect to participant satisfaction, equitable, facilitated, nurturing of critical thinking and reflection, supportive of creating a learning community, and translated into practice. These features were further categorized into five subgroups: applicability, attractiveness, functionality, learner interaction, and implementation into practice. Literature frequently cites Kirkpatrick's four-level model of evaluation and application in the review of education and training; however, few studies appeared to integrate all four levels of Kirkpatrick's model. Features were then correlated against the DCC&TN, with an encouraging connection found between these features and their inclusion within the content and structure of the DCC&TN.

Conclusions: A total of 14 key features were identified that support an effective and functional Web-based learning environment. Few studies incorporated Kirkpatrick's salient elements of the model—reaction, learning, behavior, and results—in their evaluation and clinical application. It could, therefore, be considered prudent to include Kirkpatrick's levels of training evaluation within

studies of dementia training. There were few studies that evaluated Web-based dementia education programs, with even fewer reporting evidence that Web-based training could increase staff confidence, knowledge, skills, and attitudes toward people with dementia and be sustainable over time. The DCC&TN appeared to contain the majority of key features and is one of the few programs inclusive of hospital, community, and residential care settings. The 14 key features can potentially enhance and complement future development of online training programs for health sciences education and beyond. The DCC&TN model could potentially be used as a template for future developers and evaluators of Web-based dementia training.

(J Med Internet Res 2020;22(1):e16808) doi:10.2196/16808

KEYWORDS

education; workforce; online learning; Web-based learning; distance education; dementia; nursing; facilitated learning; competency; training; network; capability; skills; person-centered

Introduction

Background

The global dementia epidemic demands a skilled and knowledgeable workforce ready to meet its related challenges [1]. For clinicians, there is an urgent need for access to education that is user friendly, affordable, and accessible with available peer support, supervision, and access to dementia champions, especially for those working in rural and remote regions. Web-based education has changed the face of learning and provides flexible, accessible, and cost-effective platforms for the delivery of education to a wide audience, regardless of their setting or location [2].

In 2015, an estimated 46.8 million people were living with dementia globally. The number of people living with dementia is expected to reach 74.7 million by 2030 and 131.5 million by 2050. In high-income countries, this number will grow by 116% between 2015 and 2050 [3]. In 2018, the estimated cost of dementia to Australia was over Aus $15 billion [4]. People with dementia occupy up to one-quarter of Australian acute-care hospital beds [1,5] and make up 52% of all residents in residential aged-care facilities [4].

The lack of professional knowledge around treatment and care options can result in delayed or hindered access to ongoing care, treatment, and support for people living with dementia [1,5]. Web-based education provides a platform for health professionals to access flexible education to improve awareness, knowledge, and skills in delivering dementia care. The high enrollment rate—almost 10,000 people from 65 countries—in the Understanding Dementia Massive Open Online Course (MOOC) [6] highlights the interest and need for quality dementia education. However, the 38% completion rate of the MOOC [6] was relatively low, which reflects the need for a more effective and responsive learning environment.

In 2007, a dementia Web-based program—the Dementia Care Competency and Training Network (DCC&TN)—was developed based on recommendations from a report commissioned by New South Wales (NSW) Health by Wylie et al [7,8]. This program aims to advance the knowledge, skills, and practice of clinicians and is facilitated by dementia champions. Our definition of a dementia champion is a "clinician who has excellent knowledge and skills in the care of the person with dementia and has a commitment to provide information and support to those undertaking the online courses." The DCC&TN delivers a Web-based learning platform that is interactive, multimodal, and facilitated. Additionally, the program includes dementia care competencies developed by Traynor et al [9,10]. These competencies are available to enhance the learner's knowledge, skills, and attitudes in the delivery of person-centered care and are freely available. Prior to the development of these competencies, Traynor and coworkers reported that no dementia competency framework existed that was applicable across care settings or levels of practice [10].

Since the inception of the DCC&TN, it has delivered high-quality dementia education and resources to over 10,000 clinicians across NSW, Australia, with an average completion rate of 78%. The program is a key training resource for NSW dementia clinical nurse consultants with a focus on a person-centered approach to delivering dementia care. The syllabus content of the DCC&TN aligns with the Australian Commission on Safety and Quality in Health Care Standards [5], with the goal of embedding these standards into clinical practice to improve the quality of life and outcomes for people living with dementia.

The literature suggests that Web-based education is a flexible [6,11-14] and cost-effective [6,15-18] medium; however, to be effective and functional it must contain features that promote learning and support knowledge translation into practice [19-21]. To our knowledge, there is limited research that has evaluated Web-based dementia education programs. An effective and functional Web-based program is defined as one that measures results, is accessible, is user friendly, and translates to clinical practice.

The DCC&TN provides a multifaceted education platform that assists clinicians in meeting the challenges of caring for the person living with dementia and in meeting professional obligations for lifelong learning. The training network is an online website comprising a content management system and a learning management system integrated into a single user experience that provides continuous membership, allowing ongoing access to resources, tools, clinical experts, and interactive forums. The library includes resources referenced within five online courses.

Objectives

This review aims to (1) determine the key features that are associated with an effective and functional Web-based education program, which is defined as one that measures results, is

accessible, is user friendly, and translates into clinical practice, and (2) determine how these features correlate with the DCC&TN.

Methods

Overview

Integrative review methodology [22] was chosen for several reasons. First, it allows for the inclusion of both experimental and nonexperimental research, so as to improve our understanding of a phenomenon. Second, integrative reviews can also combine empirical literature together with theoretical frameworks [22]. Lastly, integrative reviews address a diverse range of purposes, including defining concepts, reviewing theories or evidence, and analyzing methodological issues. The wide-ranging sampling frame, together with the diversity of purposes of integrative reviews, can assist in understanding complex concepts and theories, such as Web-based dementia care education. Hence, this review included both qualitative and quantitative studies as well as literature that looked at theories and empirical studies [22].

Study Selection

Six electronic databases—Medline, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL), AusHealth, Nursing@Ovid, and Google Scholar—were searched for studies published between 2009 and 2018. Keywords used included the following: Education, Distance, Continuing, Learning, Online, Web-Based, Internet, Dementia, Program Evaluation, Validation Studies, Outcome and Process Assessment Healthcare, Nursing, Assisted Instruction, and Facilitated (see Multimedia Appendix 1). The Critical Appraisal Skills Programme (CASP) [23] and Kirkpatrick's salient elements of the model for the evaluation of training were used to ensure quality and rigor of the analysis [18]. Search strategies were carried out using Medical Subject Headings (MeSH), relevant terms citations, and abbreviations (see Multimedia Appendix 1). Citation tracking and reference list inspections were undertaken in the search for relevant papers. Three clinical experts (KDS, AM, and KW) in dementia education conducted three stages of study selection: (1) an initial screening of titles, (2) a review of titles and abstracts, and (3) a review of the full text to identify suitable articles for inclusion. Consensus was used in the case of discrepancies.

Eligibility Criteria

Inclusion Criteria

Studies had to adhere to the following inclusion criteria in order to be included in the review:

1. Population: health personnel across different care settings or levels of practice.
2. Concept: educational interventions that included (1) Web-based learning, online learning, Internet-based education, or computer‐assisted instruction, and (2) interactive facilitated education or tutor-supported education. Studies also had to measure learner satisfaction, knowledge, skills, or behavior.
3. Context: articles from countries with similar health care provision to Australia were included.
4. Types of studies: quantitative and qualitative research papers were included. Additionally, comparison studies, literature reviews, case studies, cohort studies, systematic reviews, and randomized controlled trials were included. Studies that contained experiential and correlational designs, quasi designs, flexible learning, and Web-based learning were also included.

Exclusion Criteria

Exclusion criteria were as follows: studies that excluded information technology (IT), studies published prior to 2009, studies that were unobtainable, non-English literature, studies with unrelated relevance to the review objectives, studies with nonhealth-related context, protocol descriptions, and studies with limited reporting on translation of learning or outcomes.

Quality Appraisal, Abstraction, and Synthesis

The CASP system of appraisal [23] was adopted to evaluate the final studies for rigor, methods, credibility, and relevance. CASP is a well-utilized tool to enhance the utility of evidence-based research by health professionals. Each article was critiqued for design, methods, and study detail including aims, ethical considerations, sample population and size, interventions, and outcome measures (see Textbox 1).

Data Extraction

A standardized data extraction process to perform data extraction was used. Any discrepancies were resolved by consensus. Data extracted from each eligible study was entered into a standardized form and included the following:

1. General information: author, year of publication, and location.
2. Study characteristics: aim, study design, and ethics.
3. Sample population and size.
4. Comparative interventions.
5. Outcome measures and instruments.
6. Main findings.

A rating criteria framework was developed and agreed upon by the authors (see Textbox 1) to compare all studies so that recurring key features could be identified and subsequently applied during the review. These criteria also included questions on possible correlation of the studies with the components and characteristics of the DCC&TN.

Textbox 1. Rating criteria applied across the 46 articles included in the review.

- Correlation to the Dementia Care Competency and Training Network (DCC&TN): strong, medium, or weak
- Findings identified with evidence and in context
- Key concepts and aims
- Contribution to a wider understanding of online learning
- Lessons learned from literature and application to the DCC&TN
- Methods, omissions analysis, and validity
- Research conclusions to the objectives
- Learner satisfaction

Association of Key Features With the Dementia Care Competency and Training Network

Following the review, 14 key features were compared with the DCC&TN. Correlation with the features found in the review and the content and delivery of the DCC&TN was considered to be *strong* if eight or more features matched the DCC&TN, *medium* if between four and seven features matched the DCC&TN, and *weak* if three or fewer of the features matched the DCC&TN.

Ethics Approval and Consent to Participate

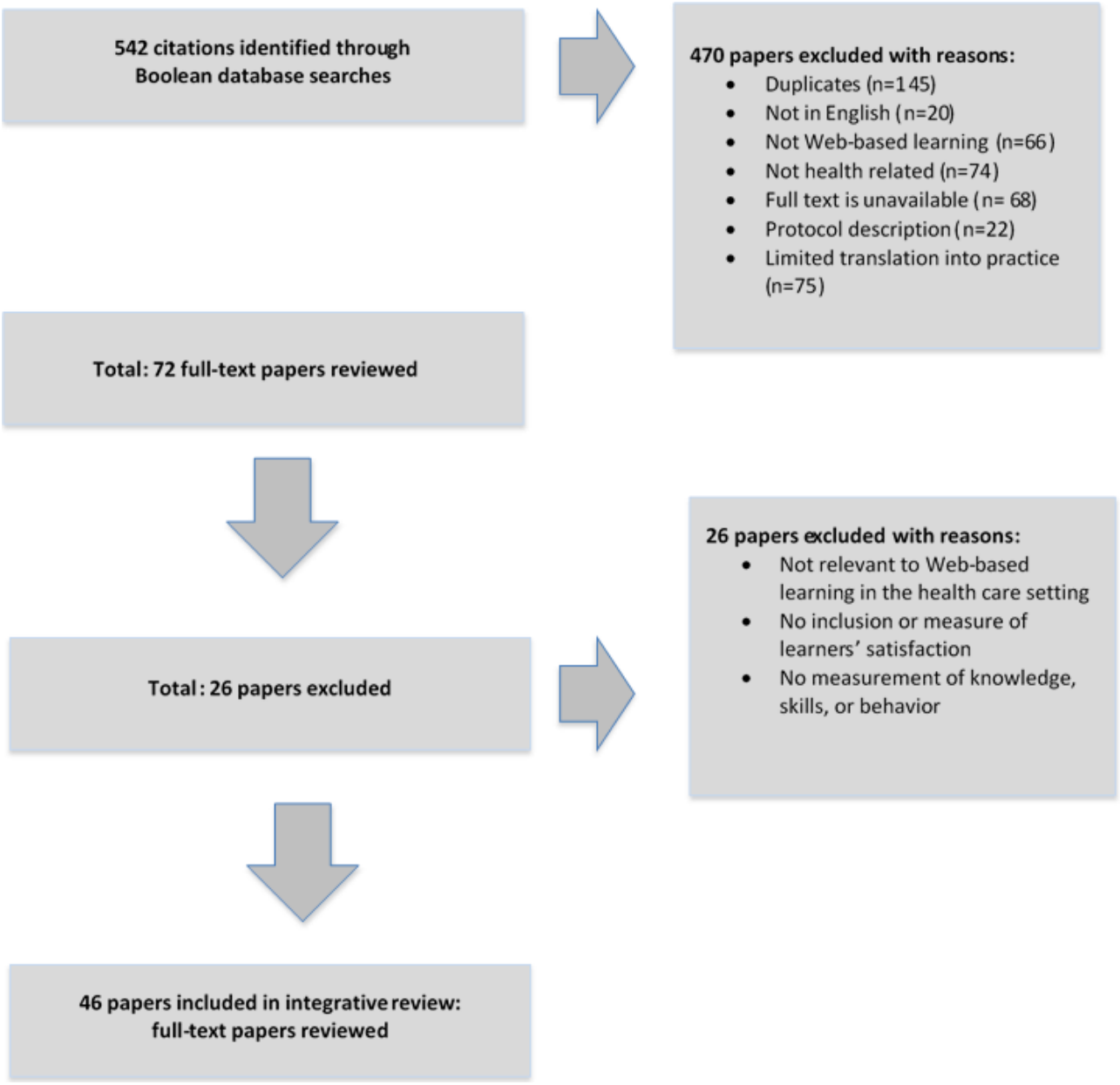
Ethics approval from a human research ethics committee was not required, as this was a systematic review of published literature.

Results

Search Outcome

There were three phases of the search process. Initially, the systematic search produced 542 citations; 470 were excluded for the following reasons: duplication; non-English; full text unavailable; relevance to objectives; context not health related; not inclusive of, or not delivered via, the Internet or not Web based or online; protocol descriptions; and limited reporting on translation of learning or outcomes (see Figure 1). Secondly, 72 full-text papers were subsequently reviewed; of these, an additional 26 papers were excluded. The third phase included the review of the remaining 46 articles, which were matched to the rating criteria in Textbox 1. The review of articles included the following study designs: 6 randomized controlled trials, 30 cross-sectional studies, and 10 literature reviews.

Figure 1. Flow diagram of the systematic review search.



Setting and Participants

Multimedia Appendix 2 displays a summary of the 46 included papers [6-56]. Most studies were conducted in the United States (13/46, 28%), followed by the United Kingdom (10/46, 22%) and Australia (8/46, 17%), with the remaining 33% (15/46) from a variety of countries. Web-based training programs were delivered to health professionals.

Data Synthesis of Key Features

The qualitative synthesis of the 46 articles was structured on what features were linked to an effective and functional Web-based learning experience and how this correlates with the DCC&TN. During data analysis, two initial observations were made. First, one seminal systematic review published in 2009 by Booth et al [14] identified five broad themes: peer communication, flexibility, support, knowledge validation, and course presentation and design. This was further supported by 11 subthemes, which provided a valuable framework for ongoing course development, as suggested by Booth et al and the literature.

Second, 6 articles referred to Kirkpatrick's salient elements of the four-level model, which are applied to determine "return on investment" and to "show [that] the business value and worth of training" was used to evaluate training and education provision [12,18,37,56]:

1. Level 1: Examines the learner's reaction to and satisfaction with the program.
2. Level 2: Assesses the extent of learning and includes knowledge, skills, confidence, and attitudes.
3. Level 3: Explores the extent to which completion of the training leads to staff behavior or practice change.
4. Level 4: Assesses the results or outcomes of training, for example, in terms of quality of patient care.

Surr and Gates [19,38], Ellis et al [12], and Scerri et al [39] refer to Kirkpatrick's salient elements of the model of reaction,

learning, behavior, results, and outcomes. Surr and Gates also reported that it is a weakness not to consider Kirkpatrick's framework within dementia education [38].

Upon further analysis of the 46 articles, 14 key features were identified and applied to what were considered effective and functional features for Web-based learning programs (see Table 1 and Multimedia Appendix 2). Some key features were used more frequently in studies, such as *self-paced* or *self-directedlearning* (n=29) and the program being interactive (n=28). Key features that were used less frequently included *cost-effectiveness* (n=10), *provides equitable engagement* (n=13), and *individualized, based on learner's profile and background* (n=15).

In total, 5 studies reported that training can increase staff confidence, knowledge, and attitudes toward those with dementia [20,21,33,39,40]. Gagnon et al [33] reported that the use of self-directed educational modules improved nurses' knowledge and skills in relation to evidence-based practice. Du et al's [21] systematic review identified learners as having a high satisfaction rate in regard to Web-based education as well as having improvement in knowledge and skills performance and enhanced self-efficacy in accomplishing nursing skills. Very few studies reported completion rates or whether learnings were translated into clinical practice. The themes that emerged from the review could be incorporated into future design elements of dementia-related Web-based learning programs. Some examples include the following:

1. Delivery of an entire learning experience that goes beyond the module and syllabus design [17,19,24].
2. Development of learning communities [13,21,31,41,45].
3. Connecting learning to improved clinical practice [15,17,26,28,38,54].
4. Learning that leads to the best possible provision of care for people living with dementia [17,21-49].

A number of papers highlight barriers faced by online learners, including lack of time, competing interests, reliability of IT, organizational support during work hours, computer access, confidence with the computer, and ability to work at their own pace [19,38,44]. It is important that these factors are considered in the development of Web-based learning. For example, issues such as lack of time, competing interests, and organizational support may potentially be addressed by content design, which includes short, sharp training measures in the way Web-based education is delivered.

It was frequently reported that learners who are engaged in Web-based studies require a level of commitment and willingness, the ability to develop self-direction, and a capacity for flexibility [14,19-37,42,43]. The authors' evaluation of the learners' feedback regarding the DCC&TN emphasized similar challenges.

Application of Key Features to the Dementia Care Competency and Training Network Program

Following the literature review, the 14 key features were cross-referenced with the syllabus content, the interactive elements of the learning platform, and the delivery environment of the DCC&TN. Table 1 shows that all features were found to correlate with the program; some features correlated strongly while others had only a weak correlation. Of the 46 studies reviewed, 57% (26/46) had a strong correlation, 28% (13/46) had a medium correlation, and 15% (7/46) had a weak correlation to the DCC&TN. Overall, the features identified in the literature, which are required for a functional and effective Web-based learning environment, are embedded in the content and structure within the learning management system of the DCC&TN.

Table 1. Thematic analysis of the literature identifying key features for effective and functional online learning.

Theme number	Key feature	Number of articles[a] that included the key feature, n	Application of key feature to the Dementia Care Competency and Training Network (DCC&TN)
1	Self-directed and self-paced	29	Learners choose when and where to engage in the courses at a time that suits their personal commitments
2	Individualized and based on learner's profile and background	15	Learners can choose courses based on their interests, competencies, and experiences
3	Interactive	28	Learner interaction occurs via real-time live chats and forums
4	Multimodal	22	Interactive and multimodal Web-based lessons are delivered, including video, interaction via the Web, and literature
5	Flexible	23	Interaction and learning is available at a time that suits the learner
6	Accessible	23	The program develops a dementia community of learners as they are enrolled in individual groups, regardless of location; the program has a user-friendly format
7	Consistency of information, repetition, and reinforcement	21	Questionnaires, feedback, surveys of satisfaction, and case discussions are embedded throughout courses and are reinforced by current literature and clinical champions
8	Cost-effective and good value for investment, both for the learner and the system	10	All courses are free and available to anyone who applies, regardless of professional background or geographical location; established education program is for use by educators and facilitators
9	Measures using questionnaires, feedback, and surveys of satisfaction	28	Various feedback mechanisms are utilized, including SurveyMonkey, learner satisfaction comments, and pre-post questionnaires
10	Provides equitable engagement	13	All courses are free and available to anyone who applies, regardless of professional background or geographical location
11	Facilitated, access to instructor, or mentored	25	Dementia facilitators support learners by providing weekly updates and encouragement, undertaking grading, and responding to individual emails for those learners who have not completed their learning milestones
12	Nurtures critical thinking and reflection	26	Reflective practice occurs via forum posts and discussion
13	Establishment of a learning community	22	Learners interact in real-time chats and forum posts, sharing their professional and personal experiences, providing case discussion, and sharing of clinical procedures and policies among their course group, thereby consolidating a learning community
14	Ability for translation into practice	17	Learners are encouraged to undertake activities or projects that demonstrate translation of learning into practice within the workplace

[a]See Multimedia Appendix 2 for individual references.

Discussion

Principal Findings

This literature review identified key features that could contribute to effective and functional Web-based learning programs. The literature frequently cited Kirkpatrick's model of evaluation and application in the review of education and training, identifying four significant elements: reaction, learning, behavior, and results. These four elements align with the expanded 14 features identified in this literature review. Additionally, the review suggests that the structure and content of the DCC&TN nurtures critical thinking within a learning community, via support and facilitation by dementia champions, and ultimately encourages translation of learning into effective person-centered practice. This has been further substantiated by the evaluation of the participant's response to feedback questionnaires during the learning experience. The DCC&TN achieves completion rates of 78% each year compared to, for example, 38% in the MOOC dementia education training program [6]. The DCC&TN has also aligned its syllabus content and resources with the Australian National Safety and Quality Health Service Standards of clinical care.

Surr and Gates [38] conducted a systematic review on effective dementia education and training for the health and social care workforce. They concluded that none of the reviews examined each of the salient elements of Kirkpatrick's four levels of

evaluation. They also found that none of the reviews combined the elements to understand the full context of key features that lead to an efficacious dementia training program in the hospital setting. This was determined to be a weakness in the current dementia literature [19]. In total, 8 studies in this review highlighted the importance of including the evaluation of staff confidence, knowledge, and attitudes toward those living with dementia [19,21-49]. Additionally, the majority of studies did not report on completion rates or whether clinical practices could be sustained over time [20,54]. The authors suggest that these are important features for translating learning into practice and should be incorporated into future evaluations. This review indicates that there is still limited evidence on the effectiveness of training in changing staff confidence, knowledge, and attitudes [37,39].

This analysis also identified a shortfall, in that much of the literature does not consider the importance of the following identified themes: training design, content or a delivery mode that is self-directed and individualized [14,38], interactivity [6,14], multimodality [6,14], flexibility [17,36,54], accessibility [17,36,54], consistency [6], and cost-effectiveness [6,17]. The findings indicate that greater depth and breadth of knowledge and education is needed to have an impact on clinicians' feelings of caring efficacy, positive attitudes, and satisfaction toward people with dementia [13,19,39,40,47,49,54].

Many of the 14 features were found across the literature in varying degrees, with Booth et al [14] identifying five similar themes and Kirkpatrick's four salient themes, further validating the 14 key features linked to an effective Web-based education environment, as follows: self-directed, individualized, interactive, multimodal, flexible, accessible, consistent, cost-effective, measurable with respect to participant satisfaction, equitable, facilitated, nurturing of critical thinking, supportive of creating a learning community, and translated into practice. The DCC&TN is able to demonstrate correlation in varying degrees to each of these features. The DCC&TN has established a *real-world* application of what is fundamental to enable translation of dementia education to the clinical coalface.

Barriers

The literature reports that Web-based learners are challenged by the following: lack of time, busy workplace, reliability of IT, limited organizational support during work hours, computer access, confidence with computers, and the ability to work at their own pace. Learners require a strong commitment to external studies, requiring a level of discipline, a willingness to develop self-direction, and a capacity for resilience [14,19-37,42,43]. It is the authors' opinion that a work-life balance needs to be addressed to encourage online learning among health professionals, their peers, and their managers.

Strengths and Limitations

Some of the excluded articles could have contributed further in emphasizing limitations for effective dementia Web-based training. The selection of papers reviewed were diverse and provided a good overview for comparison. Caution should be used when interpreting the results in order to consider outcomes that may have been reported elsewhere outside of this review.

Recommendations and Future Research

First, the DCC&TN can potentially provide an effective professional development platform and simultaneously advocate that improved outcomes for people living with dementia can be achieved with effective, functional, Web-based training programs. The ideal outcome for any Web-based dementia program is to improve the quality of life and well-being of people with dementia through the delivery of person-centered care by a skilled and knowledgeable workforce. Therefore, future developers seeking to design and develop new and innovative Web-based learning programs for dementia clinicians could be well-inspired by the exemplary DCC&TN structure and delivery mode. Second, clinicians require better access to free learning opportunities to be informed and competent in knowledge translation [57]. Our DCC&TN is provided for free to clinicians to learn about dementia person-centered care and how to apply learned knowledge to practice. Third, there are many apps that offer functions that have resulted in reducing the burden and improving health outcomes of caregivers [58]. It will be interesting to conduct further research to identify whether the same 14 features are important for training family caregivers in an online environment. Fourth, currently there is an Aged Care Royal Commission in place in Australia. The Australian Government has responded to one of the key recommendations that has come out of the Commission and has announced that it will deliver Aus $10 million for additional dementia training and support for aged-care workers and providers [59]. The results from our study can potentially be used to assist in delivering this key recommendation. Fifth, the International Organization for Standardization aims for harmonization of products and services globally. Recently, a technical committee has been set up in the area of aging societies [60]. The results from this study can potentially be used to inform standardization processes in the area of online dementia training.

Conclusions

This review identified 14 key features that are linked to deliver a functional and effective online dementia learning environment. The DCC&TN demonstrates correlation in varying degrees to each of the identified key features required for an effective and functional Web-based learning environment in that it delivers a platform that is self-directed, individualized, interactive, multimodal, flexible, accessible, consistent, and cost-effective. It is suggested that critical thinking is nurtured within a learning community supported by dementia facilitators, while encouraging translation of learning into practice.

Acknowledgments

We would like to thank Katelyn Winkworth for her support in reviewing this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of a search strategy to identify papers for this review.
[DOC File , 50 KB - jmir_v22i1e16808_app1.doc]

Multimedia Appendix 2
Summary of all included studies.
[DOC File , 280 KB - jmir_v22i1e16808_app2.doc]

References

1. Alzheimer's Australia. 2014 Jun. Dementia Care in the Acute Hospital Setting: Issues and Strategies. URL: https://www.dementia.org.au/files/Alzheimers_Australia_Numbered_Publication_40.PDF [accessed 2020-01-06]
2. Cartwright J, Franklin D, Forman D, Freegard H. Promoting collaborative dementia care via online interprofessional education. Australas J Ageing 2015 Jun;34(2):88-94. [doi: 10.1111/ajag.12106] [Medline: 24118831]
3. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. World Alzheimer Report 2015: The Global Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends. London, UK: Alzheimer's Disease International (ADI); 2015 Aug. URL: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [accessed 2020-01-06]
4. Dementia Australia. 2019 Apr. Dementia statistics. URL: https://www.dementia.org.au/statistics [accessed 2020-01-06]
5. Australian Commission on Safety and Quality in Health Care. A Better Way to Care: Safe and High-Quality Care for Patients With Cognitive Impairment (Dementia and Delirium) in Hospital—Actions for Health Service Managers. Sydney, Australia: ACSQHC; 2014. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/A-better-way-to-care-Actions-for-health-service-managers.pdf [accessed 2020-01-06]
6. Goldberg LR, Bell E, King C, O'Mara C, McInerney F, Robinson A, et al. Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. BMC Med Educ 2015 Mar 26;15:60 [FREE Full text] [doi: 10.1186/s12909-015-0344-z] [Medline: 25889490]
7. Wylie K, Alle A, Hillier G, Higgins I, Penning C, Byles J. Models of Good Practice, Nationally and Internationally, on Dementia. Specific Education for Hospital Staff in Acute Care Settings: A Review of the Literature. Newcastle, Australia: Centre for Research and Education in Ageing (CREA), University of Newcastle, in partnership with NSW Centre for Rural & Remote Mental Health; 2005.
8. Dementia Care Competency & Training Network. URL: https://dementiacare.health.nsw.gov.au/ [accessed 2020-01-06]
9. Traynor V, Cumming A, Britten N. Developing an Inter-Disciplinary Cross-Setting Dementia Care Competency Framework. Wollongong, Australia: University of Wollongong; 2015 Jan. URL: http://www.adhere.org.au/pdf/Developing_an_Inter-disciplinary_Cross-Setting_Dementia_Care_Competency_Framework.pdf [accessed 2020-01-06]
10. Traynor V, Inoue K, Crookes P. Literature review: Understanding nursing competence in dementia care. J Clin Nurs 2011 Jul;20(13-14):1948-1960. [doi: 10.1111/j.1365-2702.2010.03511.x] [Medline: 21401762]
11. Berger J, Topp R, Davis L, Jones J, Stewart L. Comparison of Web-based and face-to-face training concerning patient education within a hospital system. J Nurses Staff Dev 2009;25(3):127-132. [doi: 10.1097/nnd.0b013e3181a56ba0]
12. Ellis C, Roland D, Blair M. Professional educational interventions designed to improve knowledge and uptake of immunisation. Community Pract 2013 Jun;86(6):20-23. [Medline: 23821877]
13. Innes A, Kelly F, McCabe L. An evaluation of an online postgraduate dementia studies program. Gerontol Geriatr Educ 2012;33(4):364-382. [doi: 10.1080/02701960.2012.702166] [Medline: 23095221]
14. Booth A, Carroll C, Papaioannou D, Sutton A, Wong R. Applying findings from a systematic review of workplace-based e-learning: Implications for health information professionals. Health Info Libr J 2009 Mar;26(1):4-21 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00834.x] [Medline: 19245639]
15. Maloney S, Nicklen P, Rivers G, Foo J, Ooi YY, Reeves S, et al. A cost-effectiveness analysis of blended versus face-to-face delivery of evidence-based medicine to medical students. J Med Internet Res 2015 Jul 21;17(7):e182 [FREE Full text] [doi: 10.2196/jmir.4346] [Medline: 26197801]
16. Nguyen T. The effectiveness of online learning: Beyond no significant difference and future horizons. J Online Learn Teach 2015 Jun;11(2):309-319 [FREE Full text]
17. Riley K, Schmidt D. Does online learning click with rural nurses? A qualitative study. Aust J Rural Health 2016 Aug;24(4):265-270. [doi: 10.1111/ajr.12263] [Medline: 26689293]

18. Kirkpatrick J, Kayser Kirkpatrick W. The Kirkpatrick Four Levels™: A Fresh Look After 50 Years, 1959 - 2009. Newnan, GA: Kirkpatrick Partners; 2009 Apr. URL: http://www.kirkpatrickpartners.com/Portals/0/Resources/Kirkpatrick%20Four%20Levels%20white%20paper.pdf [accessed 2020-01-06]
19. Surr CC, Gates C, Irving D, Oyebode J, Smith SJ, Parveen S, et al. Effective dementia education and training for the health and social care workforce: A systematic review of the literature. Rev Educ Res 2017 Oct;87(5):966-1002 [FREE Full text] [doi: 10.3102/0034654317723305] [Medline: 28989194]
20. Carroll C, Booth A, Papaioannou D, Sutton A, Wong R. UK health-care professionals' experience of online learning techniques: A systematic review of qualitative data. J Contin Educ Health Prof 2009;29(4):235-241. [doi: 10.1002/chp.20041] [Medline: 19998474]
21. Du S, Liu Z, Liu S, Yin H, Xu G, Zhang H, et al. Web-based distance learning for nurse education: A systematic review. Int Nurs Rev 2013 Jun;60(2):167-177. [doi: 10.1111/inr.12015] [Medline: 23691999]
22. Whittemore R, Knafl K. The integrative review: Updated methodology. J Adv Nurs 2005 Dec;52(5):546-553 [FREE Full text] [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
23. Critical Appraisal Skills Programme (CASP). 2018. CASP checklists. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2020-01-06]
24. Shatto B, Erwin K. Moving on from Millennials: Preparing for Generation Z. J Contin Educ Nurs 2016 Jun 01;47(6):253-254. [doi: 10.3928/00220124-20160518-05] [Medline: 27232222]
25. McGowan BS, Balmer JT, Chappell K. Flipping the classroom: A data-driven model for nursing education. J Contin Educ Nurs 2014 Nov;45(11):477-478. [doi: 10.3928/00220124-20141027-11] [Medline: 25365182]
26. Ballman K, Garritano N, Beery T. Broadening the reach of standardized patients in nurse practitioner education to include the distance learner. Nurse Educ 2016;41(5):230-233. [doi: 10.1097/nne.0000000000000260]
27. Hardy S, Mushore M, Goddard L. Supporting student mental health nurses in clinical placement through virtual in-practice support (VIPS): Innovation uptake and the 'VIPS' project. Nurse Educ Today 2016 Nov;46:133-138. [doi: 10.1016/j.nedt.2016.08.033] [Medline: 27639212]
28. Bourhy H, Troupin C, Faye O, Meslin F, Abela-Ridder B, Sall AA, et al. Customized online and onsite training for rabies-control officers. Bull World Health Organ 2015 Apr 27;93(7):503-506. [doi: 10.2471/blt.14.149849]
29. Peine A, Kabino K, Spreckelsen C. Self-directed learning can outperform direct instruction in the course of a modern German medical curriculum: Results of a mixed methods trial. BMC Med Educ 2016 Jun 03;16:158 [FREE Full text] [doi: 10.1186/s12909-016-0679-0] [Medline: 27256081]
30. Law M, Rapoport MJ, Seitz D, Davidson M, Madan R, Wiens A. Evaluation of a national online educational program in geriatric psychiatry. Acad Psychiatry 2016 Dec;40(6):923-927 [FREE Full text] [doi: 10.1007/s40596-015-0377-y] [Medline: 26108395]
31. Canniford LJ, Fox-Young S. Learning and assessing competence in reflective practice: Student evaluation of the relative value of aspects of an integrated, interactive reflective practice syllabus. Collegian 2015;22(3):291-297. [doi: 10.1016/j.colegn.2014.04.003] [Medline: 26552200]
32. Boyd MR, Baliko B, Polyakova-Norwood V. Using debates to teach evidence-based practice in large online courses. J Nurs Educ 2015 Oct;54(10):578-582. [doi: 10.3928/01484834-20150916-06] [Medline: 26431518]
33. Gagnon J, Gagnon MP, Buteau RA, Azizah GM, Jetté S, Lampron A, et al. Adaptation and evaluation of online self-learning modules to teach critical appraisal and evidence-based practice in nursing: An international collaboration. Comput Inform Nurs 2015 Jul;33(7):285-294; quiz E1. [doi: 10.1097/CIN.0000000000000156] [Medline: 25978538]
34. Badiyepeymaie Jahromi Z, Mosalanejad L. Integrated method of teaching in Web Quest activity and its impact on undergraduate students' cognition and learning behaviors: A future trend in medical education. Glob J Health Sci 2015 Jan 14;7(4):249-259 [FREE Full text] [doi: 10.5539/gjhs.v7n4p249] [Medline: 25946931]
35. Ramaswamy R, Leipzig RM, Howe CL, Sauvigne K, Usiak C, Soriano RP. The portal of geriatrics online education: A 21st-century resource for teaching geriatrics. J Am Geriatr Soc 2015 Feb;63(2):335-340. [doi: 10.1111/jgs.13246] [Medline: 25644187]
36. Phillips BN, Turnbull BJ, He FX. Assessing readiness for self-directed learning within a non-traditional nursing cohort. Nurse Educ Today 2015 Mar;35(3):e1-e7. [doi: 10.1016/j.nedt.2014.12.003] [Medline: 25620290]
37. Reviriego E, Cidoncha MA, Asua J, Gagnon MP, Mateos M, Gárate L, et al. Online training course on critical appraisal for nurses: Adaptation and assessment. BMC Med Educ 2014 Jul 05;14:136 [FREE Full text] [doi: 10.1186/1472-6920-14-136] [Medline: 24996951]
38. Surr CA, Gates C. What works in delivering dementia education or training to hospital staff? A critical synthesis of the evidence. Int J Nurs Stud 2017 Oct;75:172-188. [doi: 10.1016/j.ijnurstu.2017.08.002] [Medline: 28837914]
39. Scerri A, Innes A, Scerri C. Dementia training programmes for staff working in general hospital settings: A systematic review of the literature. Aging Ment Health 2017 Aug;21(8):783-796. [doi: 10.1080/13607863.2016.1231170] [Medline: 27662075]
40. Lewis KO, Cidon MJ, Seto TL, Chen H, Mahan JD. Leveraging e-learning in medical education. Curr Probl Pediatr Adolesc Health Care 2014 Jul;44(6):150-163. [doi: 10.1016/j.cppeds.2014.01.004] [Medline: 24981664]

41. Seckman CA. Perceived sense of community, cognitive engagement, and learning outcomes among undergraduate nursing students enrolled in an Internet-based learning course. Comput Inform Nurs 2014 Oct;32(10):482-489; quiz 490. [doi: 10.1097/CIN.0000000000000076] [Medline: 24949712]
42. Arving C, Wadensten B, Johansson B. Registered nurses' thoughts on blended learning in a postgraduate course in cancer care: Content analyses of Web surveys and a focus group interview. J Cancer Educ 2014 Jun;29(2):278-283. [doi: 10.1007/s13187-013-0594-6] [Medline: 24338502]
43. Mgutshini T. Online or not? A comparison of students' experiences of an online and an on-campus class. Curationis 2013 Mar 18;36(1):E1-E7. [doi: 10.4102/curationis.v36i1.73] [Medline: 23718147]
44. Wearne SM, Greenhill J, Berryman C, Sweet L, Tietz L. An online course in clinical education: Experiences of Australian clinicians. Aust Fam Physician 2011 Dec;40(12):1001-1003 [FREE Full text]
45. Moule P, Ward R, Lockyer L. Nursing and healthcare students' experiences and use of e-learning in higher education. J Adv Nurs 2010 Dec;66(12):2785-2795. [doi: 10.1111/j.1365-2648.2010.05453.x] [Medline: 20946565]
46. Sowan AK, Jenkins LS. Designing, delivering and evaluating a distance learning nursing course responsive to students needs. Int J Med Inform 2013 Jun;82(6):553-564. [doi: 10.1016/j.ijmedinf.2013.02.004] [Medline: 23478139]
47. Kontio R, Hätönen H, Joffe G, Pitkänen A, Lahti M, Välimäki M. Impact of eLearning course on nurses' professional competence in seclusion and restraint practices: 9-month follow-up results of a randomized controlled study (ISRCTN32869544). J Psychiatr Ment Health Nurs 2013 Apr;20(5):411-418. [doi: 10.1111/j.1365-2850.2012.01933.x] [Medline: 22672441]
48. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning. J Nurses Staff Dev 2012;28(4):186-190. [doi: 10.1097/nnd.0b013e31825dfb71]
49. Surr C, Smith S, Crossland J, Robins J. Impact of a person-centred dementia care training programme on hospital staff attitudes, role efficacy and perceptions of caring for people with dementia: A repeated measures study. Int J Nurs Stud 2016 Jan;53:144-151. [doi: 10.1016/j.ijnurstu.2015.09.009] [Medline: 26421910]
50. Chao S, Chang Y, Yang S, Clark M. Development, implementation, and effects of an integrated Web-based teaching model in a nursing ethics course. Nurse Educ Today 2017 Aug;55:31-37. [doi: 10.1016/j.nedt.2017.04.011] [Medline: 28511085]
51. López Soblechero MV, González Gaya C, Hernández Ramírez JJ. A comparative study of classroom and online distance modes of official vocational education and training. PLoS One 2014;9(5):e96052 [FREE Full text] [doi: 10.1371/journal.pone.0096052] [Medline: 24788543]
52. Page J, Meehan-Andrews T, Weerakkody N, Hughes DL, Rathner JA. Student perceptions and learning outcomes of blended learning in a massive first-year core physiology for allied health subjects. Adv Physiol Educ 2017 Mar 01;41(1):44-55 [FREE Full text] [doi: 10.1152/advan.00005.2016] [Medline: 28143822]
53. Mayer B, Ring C, Muche R, Rothenbacher D, Schmidt-Strassburger U. Creating a blended learning module in an online master study programme in oncology. Educ Health (Abingdon) 2015;28(1):101-105 [FREE Full text] [doi: 10.4103/1357-6283.161951] [Medline: 26261125]
54. Ollerenshaw A, Wong Shee A, Yates M. Towards good dementia care: Awareness and uptake of an online Dementia Pathways tool for rural and regional primary health practitioners. Aust J Rural Health 2018 Apr;26(2):112-118. [doi: 10.1111/ajr.12376] [Medline: 28851026]
55. Collier E, Knifton C, Surr C. Dementia education in higher education institutions. Nurse Educ Today 2015 Jun;35(6):731-732. [doi: 10.1016/j.nedt.2015.02.013] [Medline: 25735909]
56. Hertfordshire County Council. Kirkpatrick's Evaluation Model. URL: https://www.hertfordshire.gov.uk/Media-library/Documents/Kirkpatricks-Evaluation-Model.pdf [accessed 2020-01-06]
57. Damarell RA, Tieman JJ. How do clinicians learn about knowledge translation? An investigation of current Web-based learning opportunities. JMIR Med Educ 2017 Jul 13;3(2):e12 [FREE Full text] [doi: 10.2196/mededu.7825] [Medline: 28705788]
58. Grossman MR, Zak DK, Zelinski EM. Mobile apps for caregivers of older adults: Quantitative content analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
59. Prime Minister of Australia. 2019 Nov 25. Response to Aged Care Royal Commission Interim Report URL: https://www.pm.gov.au/media/response-aged-care-royal-commission-interim-report [accessed 2020-01-06]
60. International Organization for Standardization. 2017. Ageing societies. URL: https://www.iso.org/committee/6810883.html [accessed 2020-01-06]

Abbreviations

CASP: Critical Appraisal Skills Programme
CINAHL: Cumulative Index to Nursing and Allied Health Literature
DCC&TN: Dementia Care Competency and Training Network
IT: information technology
MeSH: Medical Subject Headings
MOOC: Massive Open Online Course

NSW: New South Wales

Edited by G Eysenbach; submitted 27.10.19; peer-reviewed by W Zhang, IV George, R Chan; comments to author 16.11.19; revised version received 28.11.19; accepted 15.12.19; published 22.01.20.

Please cite as:
Moehead A, DeSouza K, Walsh K, Pit SW
A Web-Based Dementia Education Program and its Application to an Australian Web-Based Dementia Care Competency and Training Network: Integrative Systematic Review
J Med Internet Res 2020;22(1):e16808
URL: http://www.jmir.org/2020/1/e16808/
doi:10.2196/16808
PMID:32012077

Original Paper

The Impact of Advertisement Messaging on Enrollment of Young Men Who Have Sex With Men for Web-Based Research: Observational Study

Holly B Fontenot[1,2], PhD, RN, NP; Nadia N Abuelezam[1], ScD; Joshua G Rosenberger[3], PhD; David Novak[4], MSW; Kenneth H Mayer[2,5,6], MD; Gregory Zimet[7], PhD

[1]Connell School of Nursing, Boston College, Chestnut Hill, MA, United States

[2]The Fenway Institute, Boston, MA, United States

[3]Department of Biobehavioral Health, Penn State University, University Park, PA, United States

[4]DSN Consulting, LLC, Quincy, MA, United States

[5]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[6]Harvard Medical School, Boston, MA, United States

[7]Section of Adolescent Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Holly B Fontenot, PhD, RN, NP
Connell School of Nursing
Boston College
140 Commonwealth Ave
Chestnut Hill, MA, 02467
United States
Phone: 1 617 552 1846
Email: holly.fontenot@bc.edu

Abstract

Background: Recruiting young men who have sex with men (YMSM) in community settings is difficult. The use of Web-based social networks and dating apps for recruitment can be successful approaches, although little work has been done on the impact of study advertisement content on recruitment.

Objective: The aim of this study was to evaluate the effects of advertisement message content on the recruitment of YMSM (aged 18-26 years) for a Web-based focus group study, examining perspectives and preferences for a mobile app that was designed to support sexual health among YMSM.

Methods: Between March and April 2017, a recruitment campaign to promote human papillomavirus vaccination was launched on a popular social networking and dating app for YMSM, with 3 different text-based advertisement themes (technology, cancer prevention, and sexual innuendo). The campaign recruited YMSM across 3 states (Massachusetts, New York, and Pennsylvania). We examined the click-through rates, conversion rates, and enrollment rates of each of the advertisements and examined differences in views and clicks by age, state, and time of day.

Results: The sexual innuendo advertisement had the highest click rates when compared with both the technology (click rate ratio [CRR] 2.06, 95% CI 1.74-2.45) and cancer prevention (CRR 1.62, 95% CI 1.38-1.90) advertisements. The sexual innuendo advertisement also had higher study enrollment rates compared with the technology (CRR 1.90, 95% CI 1.23-2.83) and cancer prevention (CRR 2.06, 95% CI 1.37-3.13) advertisements. No differences were observed in clicks or enrollment by age, state, or time of day.

Conclusions: Our marketing campaign, targeting YMSM, was effective in recruiting participants for a qualitative study, using Web-based focus groups. The sexual innuendo advertisement was the most effective and cost-efficient advertisement of the 3 approaches trialed. Different populations need different targeted strategies for study recruitment. Researchers should work with key representatives to develop and test culturally relevant messaging and approaches that utilize current and popular technologies.

(J Med Internet Res 2020;22(1):e16027) doi:10.2196/16027

KEYWORDS

advertisement; men; sexual minorities

Introduction

Background

Improving health outcomes among young men who have sex with men (YMSM) is an important goal in the United States [1]. Numerous health disparities exist for this population [2], and health researchers are responding by actively recruiting YMSM into studies to build relevant and culturally appropriate interventions. One of the main difficulties in building relevant and culturally appropriate interventions is recruiting YMSM into feasibility studies. Issues that the youth identify as potential barriers to study enrollment include consent or assent requirements, privacy and confidentiality concerns, time, and scheduling issues [3,4]. Challenges specific to recruitment of YMSM may also include finding places to recruit sexual minorities, increasing privacy and confidentiality concerns related to minority status, and building trust, as well as a history of stigma and discrimination [5]. Recently, investigators have had success in recruiting adolescents/young adults by using Web-based approaches [6-8].

YMSM utilize Web-based resources via mobile devices or personal computers to seek health information and engage socially and sexually with other sexual minority peers [7,9]. Mobile apps for Web-based dating and sexual networks have grown in popularity over the past decade [10]. The popularity of these apps reflects YMSM's desire to have control over who they interact with and easily find sexual partners [11]. Study recruitment has been done from Web-based social networking apps and websites (eg, Facebook) [6,12] and from dating apps (eg, Grindr and Jack'd) [7,13-15]. Regardless of the Web-based venue in which recruitment is taking place, advertisements are used to attract the attention of prospective study participants. Although some work has been done to understand how the content of imagery on recruitment advertisements influences young men's engagement in research [6], little work has been done to understand how the thematic messaging of advertisements (in the absence of images) may influence engagement.

Objectives

During recruitment for a study focusing on human papillomavirus (HPV) vaccination [7], we aimed to understand how different advertisement messaging themes, delivered in a pop-up format on a popular mobile social networking and dating app for YMSM, influenced engagement with those advertisements and ultimate enrollment in the study procedures.

Methods

Human Papillomavirus Study Overview

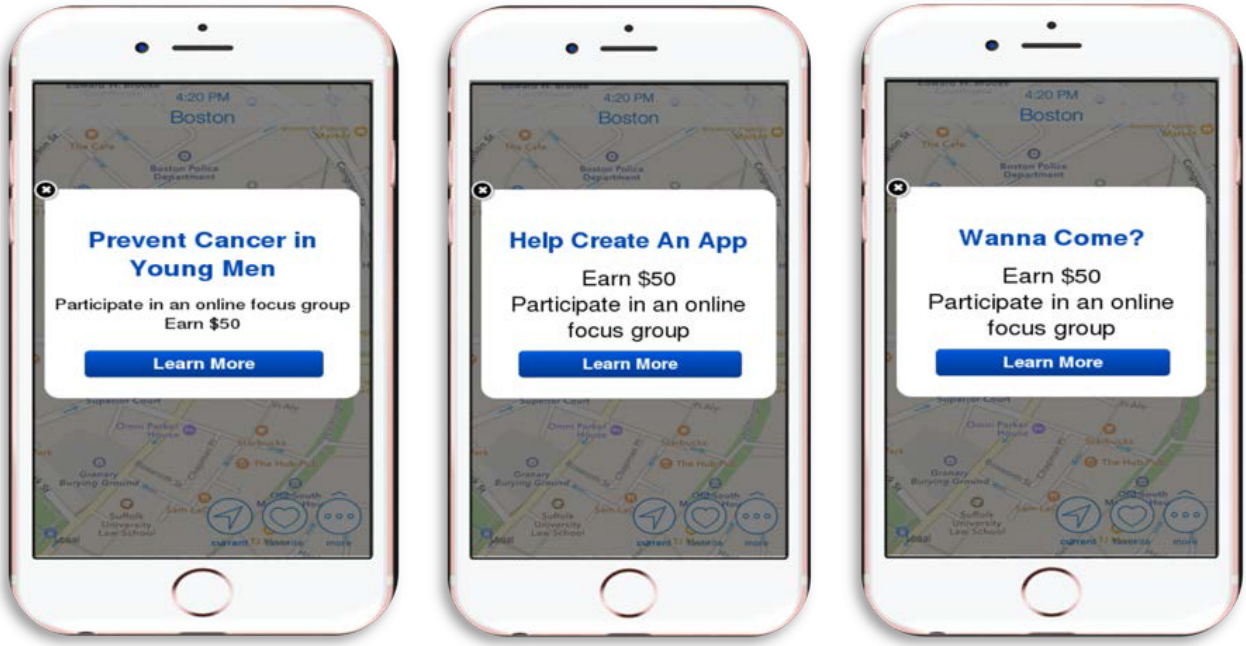
The aim of the larger study was to elicit YMSM's perspectives and preferences for an app being developed to provide information on HPV, HPV vaccination, and sexual health, as well as referral information and linkages to care. The goal was to enroll 60 YMSM across 3 northeastern states (Massachusetts, New York, and Pennsylvania) for 6 Web-based focus groups (20 YMSM per state and 2 Web-based focus groups per state). Eligibility requirements included (1) self-identification as a man who has sex with men, (2) aged 18 to 26 years, and (3) ability to read and understand English. Initial findings have been published [7].

Advertisements and Recruitment

Recruitment occurred on a popular global mobile social networking and dating app, utilized by racially diverse YMSM (app name undisclosed per company policy). A total of 3 different pop-up advertisements for study recruitment were created. Pop-up advertisements appeared on the full screen of a mobile device and were compatible for both Android and iOS. Each advertisement included a headline, text (including reference to the US $50 compensation for participation), and a button that would link potential participants to the study website. Each of the 3 advertisements had a different headline that delivered different themes; all the other text was the same for each advertisement (Figure 1). The technology-focused advertisement read as "Help Create an App" and focused on the ultimate technology/app development–related outcome of the pilot study. The cancer prevention–focused advertisement read as "Prevent Cancer in Young Men," referencing the HPV prevention/vaccination outcome of the larger study. The sexual innuendo–focused advertisement read as "Wanna come?", referencing an individual's attendance for Web-based focus groups (with sexual innuendo, as *come* spelled as *cum* is slang for ejaculate/seminal fluid). All advertisements were vetted by key representatives from the population of interest during development.

All advertisements were displayed on the dating app at the same rate during the same recruitment time frame (1 week between March 27 and April 4, 2017) geolocated to 3 different states. Our study marketing campaign comprised 25% of the market shares of all advertisements posted to the dating app during that week. The parent study and all advertisements were approved by Fenway Health Institutional Review Board.

Figure 1. A total of 3 study advertisements with different text-based themes.



Project Enrollment

Users of the dating app, who clicked on the link provided on the advertisement, were routed to the study Web page. This Web page included information about the Web-based focus group study, an eligibility survey, and a Web-based consent form. Men who completed the eligibility survey and were eligible to participate were provided with electronic consent forms and were asked to indicate their availability to participate in a 1-hour Web-based focus group. Availability was assessed with a question that provided 3 different dates and times on a drop-down list (the answer choice allowed multiple selections). Dates and times differed depending on the state in which the participant was residing. Participants were given 3 options for scheduling; thereafter, study investigators chose the 2 dates that worked best for the majority of eligible participants. Participants were then emailed invitations to participate for a scheduled focus group.

Measures

Measures included number of unique views for each advertisement, unique clicks for each advertisement, eligibility survey initiations, eligibility survey completions, men eligible, and men who provided informed consent. From these quantities, unique click-through rates were calculated, representing the number of unique clicks per 100 person views of the advertisement. Conversion rate, representing the number of completions of the eligibility survey per 100 person views of the advertisement and enrollment rate, representing the number of men found eligible and providing consent per 100 person views, were also calculated. In addition, using information on the costs for each of the ads, a calculated cost per enrollment, representing the total cost for the advertisement per enrolled (eligible and consented) participant, was captured.

Data Analysis

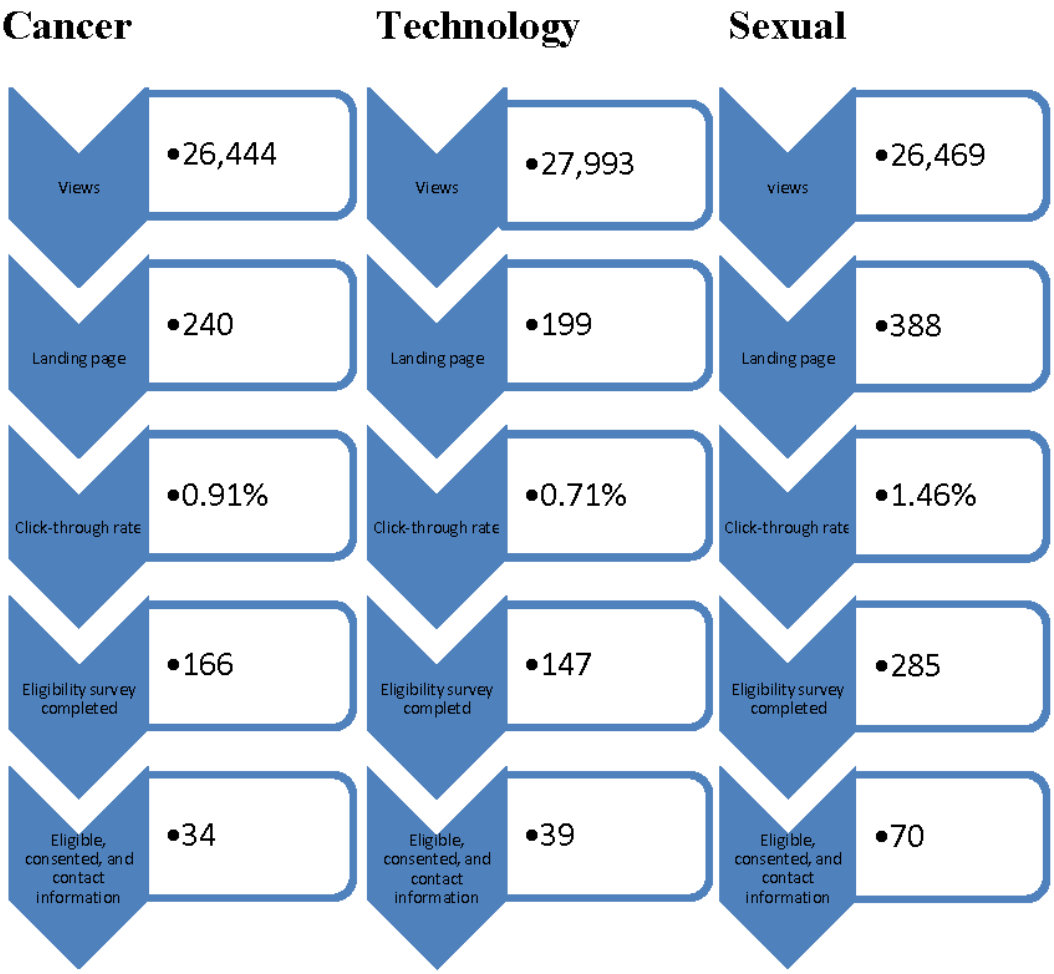
Descriptive statistics for each of the advertisement messages, using information on the state where the advertisement was displayed, and the age of survey takers were calculated. Chi-square tests were used to compare men from different states and different age categories across advertisement types. Rate ratios were calculated (unique click-through rate, conversion rate, and enrollment rate), along with their 95% confidence intervals using conditional maximum likelihood estimates from contingency tables, using the technology-focused advertisement as the reference for the analyses. Analyses were conducted using SAS software version 9.4 (SAS Institute Inc).

Results

Overview

Across all advertisements, the campaign reached 80,906 potential participants. The technology advertisement received 27,993/80,906 (34.59%) unique views, the cancer advertisement received 26,444/80,906 (32.68%) unique views, and the sexual innuendo advertisement received 26,469/80,906 (32.71%) unique views (Figure 2). There were 827 unique clicks on the advertisements, resulting in an overall unique click rate of 1.02% (827/80,906). Of those who clicked on the advertisements, 598 participants completed the eligibility survey, resulting in an overall conversion rate of 0.74% (598/80,906). A total of 143 men were eligible for the study, and they provided informed consent, resulting in an overall enrollment rate of 0.18% (143/80,906).

Figure 2. Sexual innuendo advertisement outperformed other advertisements.



Advertisement-Specific Results

The theme of the advertisement-affected response rates (Table 1). Click rates were higher for the advertisements that included a sexual innuendo theme (click rate ratio [CRR] 2.06, 95% CI 1.74-2.45) or a cancer theme (CRR 1.28, 95% CI 1.06-1.54) than for the advertisement that had a technology theme (Table 1).

The advertisement that included a sexual innuendo theme also had a higher conversion rate (CRR 2.10, 95% CI 1.73-2.56) and a higher enrollment rate (CRR 1.90, 95% CI 1.23-2.83) than the advertisement that had the technology theme. Cost per enrollment was lowest for the sexual innuendo advertisement at US $21.43 per enrollment and most expensive for the technology advertisement at US $51.72 per enrollment.

Participants' Characteristics

Age, state, and time of day the advertisement was viewed were collected for men who completed the eligibility survey after clicking through the advertisement (N=598; Table 2).

More than 50% of all men were in the eligible age range for the study (18-26 years of age). The majority responded from New York (339/598, 56.7%) followed by Pennsylvania (152/598, 25.4%) and Massachusetts (107/598, 17.9%). More men in New York responded to the technology advertisement than in Pennsylvania (27.7% [94/339] vs 20.4% [31/152]), whereas more men in Massachusetts responded to the sexual innuendo advertisement than men in New York (54.2% [58/107] vs 44.3% [150/339]), although none of these differences were statistically significant. We found no differences in advertisement selection on the basis of time of day when the advertisements were viewed.

Table 1. Unique click-through, conversion, and enrollment rates are presented for each of the advertisement themes. Cost per enrollment is also reported for each advertisement theme.

Advertisement type	Unique click-through rate (clicks per 100 person views)		Conversion rate (to complete eligibility survey)		Enrollment rate (eligible, consented, and contact info)		Cost per enrollment (US $)
	Mean	CRR[a] (95% CI)	Mean	CRR (95% CI)	Mean	CRR (95% CI)	
Sexy advertisement versus tech advertisement							
Tech	0.71	Reference	0.54	Reference	0.14	Reference	51.72
Cancer	0.91	1.28 (1.06-1.54)	0.63	1.16 (0.93-1.45)	0.12	0.92 (0.58-1.46)	44.12
Sexy	1.47	2.06 (1.74-2.45)	1.13	2.10 (1.73-2.56)	0.26	1.90 (1.23-2.83)	21.43
Sexy advertisement versus cancer advertisement							
Cancer	0.91	Reference	0.63	Reference	0.12	Reference	44.12
Sexy	1.47	1.62 (1.38-1.90)	1.13	1.81 (1.50-2.19)	0.26	2.06 (1.37-3.13)	21.43

[a]CRR: click rate ratio.

Table 2. Characteristics (state of residence, age, and time of day of click) for men who completed the eligibility survey by advertisement theme.

Indicator	Advertisement type			Total (N=598)
	Tech (n=147)	Cancer (n=166)	Sexy (n=285)	
State				
Massachusetts	22 (20.6)	27 (25.2)	58 (54.2)	107 (17.9)
New York	94 (27.7)	95 (28.0)	150 (44.3)	339 (56.7)
Pennsylvania	31 (20.4)	44 (29.0)	77 (50.7)	152 (25.4)
Age (years)				
18-26	88 (25.0)	100 (28.4)	88 (25.0)	352 (59.3)
27+	58 (24.0)	64 (26.5)	58 (24.0)	242 (40.7)
Time of day				
AM	71 (23.8)	87 (29.1)	141 (47.2)	299 (50.0)
PM	76 (25.4)	79 (26.4)	144 (48.2)	299 (50.0)

Recruitment Results

Of the 598 men who responded to the eligibility survey, 143/598 (23.9%) men met all the eligibility criteria and provided informed consent. Of these, 48 YMSM (13 to 19 per state), with a mean age of 23.4 years, participated in the Web-based focus groups that made up the larger study. Of note, this recruitment strategy was successful in enrolling YMSM of diverse racial and ethnic backgrounds; 70% participants reported race as black, 12% white, 4% Asian, 8% more than one race, and 6% other and 22% reported Hispanic ethnicity [7].

Discussion

Principal Findings

Our marketing campaign on 1 popular social media app that is oriented to YMSM seeking social and sexual relationships was effective in recruiting for Web-based focus groups. The campaign was rapid (approximately 1 week) and effective (overall conversion rate of 0.74% [598/80,906], leading to 143 eligible and consented individuals), and the enrollment of a racially and ethnic diverse sample of YMSM (n=48) for a qualitative study focused on HPV. The sexual innuendo theme was the most effective text-based advertisement out of the 3 text-based advertisement themes trialed.

Although all advertisements had the same market share and views, the sexual innuendo advertisement outperformed the other advertisements. The sexual innuendo advertisement had, approximately 2 times, the individuals click to the landing page, complete the eligibility survey, and report being eligible, as well as provide consent and contact information for further engagement (Figure 2). This advertisement was also the most cost effective. During the subsequent Web-based focus groups, YMSM noted that fun and sexy advertisements were a facilitator to research participation. Other researchers have also reported successful recruitment of YMSM for HPV- [6,8,12] or HIV-focused [13,16] studies on social media, nationally and internationally. Reiter et al [6] tested image-based advertisements on Facebook and reported a similar overall conversion rate of 0.66% (our study: 0.74%). They reported higher conversion rates for advertisements that had images of a couple, and advertisements with text mentioning sexually transmitted diseases had higher click-through rates as compared with advertisements with text mentioning cancer [6]. Martinez et al [16] found Facebook to also be effective, and these

researchers also reported success in recruiting Latino gay couples. Finally, Buckingham et al [13] found recruitment on Grindr (an app very similar to the app used for recruitment in this study) the most effective as compared with recruitment on other social media platforms. These researchers also reported Grindr recruitment to be the most rapid, yielding a large number of potential participants in the shorter period of time [13].

Adolescents and young adults are using websites and apps to engage socially and to seek health information [17,18], and creative marketing campaigns for research is becoming an increasingly important strategy for engagement [19]. Recruitment of YMSM and other marginalized youth for research may be challenging; however, increasingly, success has been noted with recent recruitment strategies on the Web [6,7,13,16].

Strengths and Limitations

Our results should be considered in terms of our study limitations. Our recruitment strategy (utilizing a dating app) may not be reflective of all YMSM who could benefit from HPV vaccination or who may be interested in research. YMSM who accessed the app generally did so to meet sexual partners, so they might have been more responsive to sexually oriented advertisements. Therefore, the sexual innuendo advertisement was potentially the best advertisement targeted to the audience. In addition, it is possible that YMSM responded positively to this advertisement, as it was uniquely phrased as an invitation or a question, and this phrasing may tap into young adults' desires to engage in new or perceived riskier (because of the sexual innuendo) opportunities. Future research in different Web-based environments may find different results. We were recruiting for a 1-time/1-hour Web-based focus group. Recruitment for longitudinal studies may be more difficult; however, our findings support that the youth will re-engage to participate in qualitative studies on the Web (Web-based focus groups), after the initial recruitment eligibility survey. We also offered compensation of US $50. Recruitment may be more difficult with lower remuneration, especially for longitudinal studies. In addition, the font size for compensation on the cancer advertisement was slightly smaller than the other 2 advertisements. This potentially could have had an effect on the findings. Another limitation is that this study focused on 3 large northeastern US cities. As a result, the findings and preferences noted here may not apply to recruitment of YMSM participants in small cities or in other geographic locations.

Overall, we were successful in recruiting a diverse sample of YMSM for qualitative research, despite not screening for ethnicity and race in the initial eligibility survey. Our results point to the importance of considering not just the platforms and locations used for recruitment (eg, Facebook, dating apps, and community settings) but also the phrasing and content of recruitment advertising.

Conclusions

Different populations need different targeted strategies for recruitment, including study advertisement messaging. We were successful with a technology-driven, effective, and rapid approach to recruit a diverse sample of YMSM for a qualitative study, Web-based focus groups. The text-based advertisement with a sexual innuendo theme was most effective. Researchers should collaborate with representatives from their population of interest to develop effective and culturally relevant study recruitment strategies that capitalize on current and popular technologies.

Acknowledgments

This work was supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases [1R21AI130447-01] awarded to HBF.

Conflicts of Interest

GZ received honorarium from Sanofi Pasteur for his work on the Adolescent Immunization initiative and an honorarium and travel support from Merck for a presentation at an HPV vaccine symposium. Other authors neither have conflicts of interest to report nor competing financial interests.

References

1. HealthyPeople. Lesbian, Gay, Bisexual, and Transgender Health URL: https://www.healthypeople.gov/2020/topics-objectives/topic/lesbian-gay-bisexual-and-transgender-health [accessed 2019-07-11]
2. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ 2018 Jun 15;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
3. Shaw SM, Caldwell LL, Kleiber DA. Boredom, stress and social control in the daily activities of adolescents. J Leisure Res 1996;28(4):274-292. [doi: 10.1080/00222216.1996.11949776]
4. Fredricks JA. Extracurricular participation and academic outcomes: testing the over-scheduling hypothesis. J Youth Adolesc 2012 Mar;41(3):295-306. [doi: 10.1007/s10964-011-9704-0] [Medline: 21814871]
5. Prescott TL, Phillips GI, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching adolescent gay, bisexual, and queer men online: development and refinement of a national recruitment strategy. J Med Internet Res 2016 Aug 4;18(8):e200 [FREE Full text] [doi: 10.2196/jmir.5602] [Medline: 27492781]

6. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017 Jul 2;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
7. Fontenot HB, Rosenberger JG, McNair KT, Mayer KH, Zimet G. Perspectives and preferences for a mobile health tool designed to facilitate HPV vaccination among young men who have sex with men. Hum Vaccin Immunother 2019;15(7-8):1815-1823. [doi: 10.1080/21645515.2019.1568156] [Medline: 30625049]
8. Gerend MA, Madkins K, Phillips G, Mustanski B. Predictors of human papillomavirus vaccination among young men who have sex with men. Sex Transm Dis 2016 Mar;43(3):185-191 [FREE Full text] [doi: 10.1097/OLQ.0000000000000408] [Medline: 26859806]
9. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and feasibility of using established geosocial and sexual networking mobile applications to promote HIV and STD testing among men who have sex with men. AIDS Behav 2015 Mar;19(3):543-552 [FREE Full text] [doi: 10.1007/s10461-014-0942-5] [Medline: 25381563]
10. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup app use, sexual behavior, and sexual health among adolescent men who have sex with men in the United States. J Adolesc Health 2018 Jun;62(6):708-715 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.01.001] [Medline: 29784114]
11. Miller B. 'They're the modern-day gay bar': Exploring the uses and gratifications of social networks for men who have sex with men. Comput Human Behav 2015;51:476-482. [doi: 10.1016/j.chb.2015.05.023]
12. Das R, Machalek DA, Molesworth EG, Garland SM. Using Facebook to recruit young Australian men into a cross-sectional human papillomavirus study. J Med Internet Res 2017 Nov 17;19(11):e389 [FREE Full text] [doi: 10.2196/jmir.8739] [Medline: 29150417]
13. Buckingham L, Becher J, Voytek CD, Fiore D, Dunbar D, Davis-Vogel A, et al. Going social: success in online recruitment of men who have sex with men for prevention HIV vaccine research. Vaccine 2017 Jun 14;35(27):3498-3505 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.002] [Medline: 28526330]
14. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: 10.1007/s10461-012-0277-z] [Medline: 22851153]
15. Pines HA, Gorbach PM, Weiss RE, Hess K, Murphy R, Saunders T, et al. Acceptability of potential rectal microbicide delivery systems for HIV prevention: a randomized crossover trial. AIDS Behav 2013 Mar;17(3):1002-1015 [FREE Full text] [doi: 10.1007/s10461-012-0358-z] [Medline: 23114512]
16. Martinez O, Wu E, Shultz AZ, Capote J, Rios JL, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 May 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
17. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: a national survey. Media Commun 2016;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
18. Anderson M, Jiang J. Public Services Alliance. 2018. Teens, Social Media & Technology 2018 URL: http://publicservicesalliance.org/wp-content/uploads/2018/06/Teens-Social-Media-Technology-2018-PEW.pdf [accessed 2019-10-29]
19. Ashley C, Tuten T. Creative strategies in social media marketing: an exploratory study of branded social content and consumer engagement. Psychol Mark 2015;32(1):15-27. [doi: 10.1002/mar.20761]

Abbreviations

CRR: click rate ratio
HPV: human papillomavirus
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 27.08.19; peer-reviewed by M Argenyi, M Barratt; comments to author 08.10.19; revised version received 18.10.19; accepted 20.10.19; published 13.01.20.

Please cite as:
Fontenot HB, Abuelezam NN, Rosenberger JG, Novak D, Mayer KH, Zimet G
The Impact of Advertisement Messaging on Enrollment of Young Men Who Have Sex With Men for Web-Based Research: Observational Study
J Med Internet Res 2020;22(1):e16027
URL: https://www.jmir.org/2020/1/e16027
doi:10.2196/16027
PMID:31929103

Original Paper

You Get What You Pay for on Health Care Question and Answer Platforms: Nonparticipant Observational Study

Fatemeh Ameri[1], PhD; Kathleen Keeling[1], PhD; Reza Salehnejad[1], PhD

Alliance Manchester Business School, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Fatemeh Ameri, PhD
Alliance Manchester Business School
The University of Manchester
AMBS New Building
Booth Street West
Manchester, M15 6PB
United Kingdom
Phone: 44 161 275 6460
Fax: 44 161 275 6460
Email: fatemeh.ameri@manchester.ac.uk

Abstract

Background: Seeking health information on the internet is very popular despite the debatable ability of lay users to evaluate the quality of health information and uneven quality of information available on the Web. Consulting the internet for health information is pervasive, particularly when other sources are inaccessible because of time, distance, and money constraints or when sensitive or embarrassing questions are to be explored. Question and answer (Q&A) platforms are Web-based services that provide personalized health advice upon the information seekers' request. However, it is not clear how the quality of health advices is ensured on these platforms.

Objective: The objective of this study was to identify how platform design impacts the quality of Web-based health advices and equal access to health information on the internet.

Methods: A total of 900 Q&As were collected from 9 Q&A platforms with different design features. Data on the design features for each platform were generated. Paid physicians evaluated the data to quantify the quality of health advices. Guided by the literature, the design features that affected information quality were identified and recorded for each Q&A platform. The least absolute shrinkage and selection operator and unbiased regression tree methods were used for the analysis.

Results: Q&A platform design and health advice quality were related. Expertise of information providers (beta=.48; *P*=.001), financial incentive (beta=.4; *P*=.001), external reputation (beta=.28; *P*=.002), and question quality (beta=.12; *P*=.001) best predicted health advice quality. Virtual incentive, Web 2.0 mechanisms, and reputation systems were not associated with health advice quality.

Conclusions: Access to high-quality health advices on the internet is unequal and skewed toward high-income and high-literacy groups. However, there are possibilities to generate high-quality health advices for free.

(J Med Internet Res 2020;22(1):e13534) doi:10.2196/13534

KEYWORDS

internet health information; health information; health literacy; eHealth; information literacy; health care access

Introduction

Background

Common drivers of the popular Web-based health care information seeking are serious health needs and inaccessibility of other sources in traditional settings [1] because of time, distance, and financial constraints or the value of a sense of control and empowerment or anonymity [2-5].

With resonance to the idea of *frugal innovations* in health care [6], so that "more can be done for less for many more people, globally" [7], the internet is seen as a low-cost and convenient way of accessing health information and care that could help in reducing the burden on health care systems. Bhatti et al

recommend assessing affordability, adaptability, and accessibility [7]. For accessibility and adaptability, the pervasive adoption of mobile phones helps inclusion by providing wider opportunities for access. However, equal opportunity of access to the internet does not guarantee equal access to high-quality health information, which is variable across health topic areas [8]. Moreover, people find it difficult to find relevant information when they are using search engines to find answers for their question [9-11].

The quality of health information available on the Web has long been a matter of concern [12,13]. Several top-down initiatives have been developed to monitor quality, including MedPICS Certification and Rating of Trustworthy Health Information by Health on the Net Foundation (a European Union project for certification and rating of trustworthy and assessed health information on the internet) and Information Standard (an accreditation system established by UK health departments aimed at filtering unreliable Web-based health information). The success of such measures is bound to be limited. The rise of Web 2.0 (the second generation of the Web that is interactive and dynamic) has increased the complexity of Web-based health information provision [14,15]. The diversity of Web-based health information sources and the ongoing debate regarding Web-based health information quality question the effectiveness of top-down approaches.

Furthermore, accurate information can be taken out of context and cause harm. Users may concurrently pursue multiple medications following advices they receive on the Web, with adverse consequences. The use of search engines to locate relevant information can be challenging and requires experience [16]. Question and answer (Q&A) health care information platforms are Web-based services where information seekers may request personalized health advice. This can help alleviate the individual context or need issue, but the quality remains a matter of concern.

A threat to accessibility is that not all health care Q&A sites are free to use. The paradox is that advertisement-supported platforms where health information is free can often be perceived as less credible by users [17].

As a solution to the information quality issue, information management literature recognizes the relationship between platform design features and the quality of information generated on the platform. Q&A platforms embody technical and social choices that can impact user activities and the quality of Q&As [18], including eligibility constraints such as medical expertise for answering, tracking user contributions, reputation systems, revealing identities versus anonymity, and, crucially, cost of participation. It is unclear and under-researched whether these design features are effective in health domains. The common approach of relying on user feedback to rate information quality [19] is questionable in health domains, as lay users may lack the knowledge to evaluate health information [10,20]. Such concerns raise questions about the effect of platform design on health information quality in Q&A platforms.

Research on platform design features that promote the generation of high-quality advices is scarce in health care. This study reviews the literature to create a thorough list of design features that may impact information quality and examines their relationships with health advice quality on Q&A platforms. A unique contribution of this study is merging 3 levels of design features related to quality and examining their interactive effects on information quality in health Q&As. A further contribution is the research design and dataset. Using actual Q&As from Q&A websites and applying a nonparticipatory data collection method, where the observer does not participate in the social setting, improve the precision of the research results [21]. The research employs least absolute shrinkage and selection operator (LASSO) to deal with the initial large number of variables and uses unbiased regression tree to investigate possible interactions among design features. The approach offers valuable policy insights and recommendations for how to design Q&A platforms to maximize information quality.

Model Development

Few studies focus on health information quality. We review major research streams on the effect of Web-based platform design features on information quality, classifying them into 3 categories: (1) firm level, (2) platform level, and (3) user level. Table 1 summarizes these research strands and defines the focus of this research. The main bodies of literature examined were information management, health services, designing Web-based platforms. The search began in Google scholar using the keywords “online health information quality.” From these results, the search was expanded by other keywords: “online information quality,” “information quality,” “answer quality,” “information sharing quality,” “knowledge sharing quality,” and “quality of online contribution.” Backward and forward search from highly relevant papers also expanded the search until it was determined that no new information was being uncovered. In the empirical section, we will test the relationship between these 3 levels of design features and health care advice quality.

Table 1. Model development summary and deficiencies filled by this study.

Level of analysis and factors	Studies	Focus of previous research	Focus of this study
Firm level			
Revenue model (advertisement, transaction, membership fee)	Harper et al [22]; Enders et al [23]	Comparison of information quality in fee-based and free platforms; the role of revenue model in social networking sites	Effect of revenue model
Platform level			
Incentives			
Financial	Chen et al [24]; Harper et al [22]; Wang et al [25]; Hsieh et al [26]	Financial motivation for knowledge sharing	Effects of different types of financial incentives
Nonfinancial	Chang and Chuang [27]; Chen et al [24]; Khansa et al [28]	Nonfinancial motivation for knowledge sharing, for example, altruism, social recognition, and social interaction	Effect of nonfinancial incentives in the form of points or credits
Reputation (internal, external reputation)	Hung et al [29]; Chang and Chuang [27]; Tausczik and Pennebaker [30]	Comparing the effect of reputation and financial incentives; effect of external reputation	Effect of reputation
Web 2.0 mechanisms	Khansa et al [28]	Effect of information technology–enabled incentives on knowledge sharing behavior	Effects of formal and informal mechanisms
User level			
Users knowledge background (medical certification, expertise)	Reavley et al [31]; Giles [32]; Clauson et al [33]; Harper et al [22]; Oh [34]	Comparability of information quality provided by experts and crowd sourcing process	Effect of expertise
Question quality	Hsieh and Counts [35]; Hsieh et al [26]	Variation of answer quality based on question type	Effect of question quality
Across levels			
All the above factors	N/A[a]	N/A	The interaction and interplay of all design features

[a]N/A: not applicable.

Firm-Level Design Features

Designing a revenue model is an essential part of platform design. Revenue can be generated from advertisements, membership fees, trading information, selling information to third parties, or a combination of any of these. Information economics theory suggests the revenue model choice may affect information quality. The key revenue driver for the advertisement-based revenue model is the number of users. However, maximizing Q&A numbers without regard to quality could lead to poor information generation [24,36]. Enders et al [23] argue that platforms earning revenue from information transactions attempt to maximize user willingness to pay by offering high-value information. From the user perspective, health information on advertisement-supported websites is often perceived as less credible [17], but this is moderated by context, such as whether websites might be expected or not to be allied to advertising [37]. The situation may be more nuanced than suggested by information economics.

Platform-Level Design Features

Provision of quality information is costly. A rational participant will weigh the time and effort costs of providing information against its expected rewards. Adequate financial or social rewards are critical for ensuring provision of high-quality information [22,24,25,27]. An understanding of how different types of rewards, financial versus intrinsic, impact information quality is critical for equality of access to quality health information. For poor populations to benefit, the quality of health information on platforms where no fee is charged needs to be high.

Financial Incentive

Empirical studies on the effect of financial incentives on information quality present no clear results. For Google Answers (a platform using financial incentives), Chen et al found that higher financial incentives led to longer but not better answers [24], whereas Harper et al found Google Answers to have higher answer quality information compared with free Q&A sites (eg, Yahoo Answers and Live QnA) [22]. In an experimental study of product reviews, Wang et al [25] found no significant quality differences between reviews by paid and unpaid reviewers.

Nonfinancial Incentive

Several studies suggest the efficacy of nonfinancial incentives such as virtual points and credit to leverage motivations such as recognition or reputation building to encourage participation [24,27,29]. Participation is encouraged through reputation

systems and virtual points or Web 2.0 mechanisms, such as voting or following. Oh and Syn [38] found different motivations associated with particular answering strategies. The association of these mechanisms to the quality of users' contributions is under-researched.

Intrinsic motivations can motivate information providers to share their information. In social health Q&A, empathy with others who are going through similar pain and stress is a strong motivator. Oh [34] found altruism was the most influential factor for participation of health question answerers in social health Q&As.

A platform may encourage users to reveal their real identity or sign up using their social media account such Facebook or LinkedIn. This extends the reputation of the users to outside the platform and gives them higher incentives to co-operate, as they want to protect or promote their *external reputation* [30].

User-Level Design Features

Expertise of the Information Provider

The role of information provider expertise is another contested question. For some, the quality of user-generated information in websites such as Wikipedia is comparable with, or even better than, the quality of information provided by experts in centrally controlled Web and printed sources [32]. In contrast, other studies found the opposite [31,33,39].

Quality of the Query

Shah et al [40] argue for a connection between an expressed information need and information quality provided. Indeed, Chen et al [24] indicate information providers are discouraged by insufficient numbers of high-quality questions on an information platform. If information seekers post nonserious or trivial questions, or are allowed to post unrelated issues, they automatically waste valuable time and the attention of the potential answerers [35]. Relatedly, information seekers can themselves increase answer information quality by increasing question quality [22].

Across Levels: Combination of Design Features

Each Web-based information platform contains a mixed set of design features. It is unlikely that a single design feature uniquely shapes information quality. From both a theoretical and practical perspective, it is critical to investigate the complementary role of design features. There is some evidence of such interactions for select website features [26], but to the best of our knowledge, there is no research investigating design features associated with quality at a comprehensive level.

Methods

Research Design

On July 1, 2014, we used 4 search engines (Google [41], Yahoo! [42], Bing [43], and Ask.com [44]) to search the terms "question and answer sites or platform or website." Our search was in English but not confined to any particular country. We chose the top 200 search results and obtained a total of 40 Q&A platforms. We excluded the nonhealth Q&A platforms and analyzed them to identify a set whose design features contained the variation of design features described above. In the case of similarity of mechanisms of platforms, popularity of the platform both among the internet users and in the literature made a platform more favorable for selection. This resulted in the choice of 9 platforms: AllExperts, AnswerBag, ChaCha, Google Answers, JustAnswer, Mahalo Answers, Quora, WebMD, and Yahoo Answers. Although Google Answers and Mahalo are no longer active, they are included in our sample because they represent design features not embodied in active platforms (see Multimedia Appendix 1 for description of the design features of the nominated platforms). One of the research aims was looking at how different incentives and user knowledge might affect the quality of answers; therefore, the unit of analysis was *1 question and 1 answer*, which corresponded with the analytic strategy of this exploratory study. In cases where a conversation took place between a single questioner and responder, the whole thread was evaluated. In addition, *1 question and 1 answer* were consistent with platforms that do not allow more than 1 answer.

Recruiting human raters to assess information quality is a common practice in the literature [24,26,45]. We recruited National Health Service–certified physicians and trained them (July 2014) to increase the validity and reliability of the quality evaluation. Following the random selection of 100 Q&As from each Q&A platform over the period of 6 weeks (900 in total) in July and August 2014, 2 physicians conducted the evaluations, each of them rated half of the data plus 10% to allow an inter-rater reliability check (see Multimedia Appendix 2 for details of the data selection and rating process). A first layer of anonymity was provided by the coders being *blind* to the name of the Q&A website where the question was asked and any other attributes associated to the Q&As. There is little risk to individual anonymity, as this level of detail is removed during the coding process. The research was conducted under the research ethics rules and policies of the University of Manchester, United Kingdom.

Dependent Variables

Many attempts have been made to identify the criteria to evaluate Web-based health information [15,46]. Included in these is the extensive literature review by Oh et al [47] that developed the following criteria for measuring health advice quality in Q&A platforms: accuracy, completeness, relevance, objectivity, readability, and source credibility. See Table 2 for definitions. These criteria formed the basis for our research based on the relevance to the study [47] and from medical point of view.

The 2 assessors rated the answers on each measure (5-point Likert scale, with 5=very high and 1=very low quality). We conducted principal component analysis (PCA) to make an index of answer quality from the primary measures, and 1 factor with an eigenvalue >1.0 was extracted from the PCA (see Multimedia Appendix 3 for PCA details). Thus, we built a single composite measure by averaging the sum of the individual criteria ratings.

Table 2. Health answer quality measures and definitions.

Criteria	Explanation
Accuracy	The answer provides correct information, that is, degree of concordance of the information provided with the best evidence or with generally accepted medical practice
Completeness	The answer includes all key points
Relevance	The answer is relevant to the question
Objectivity	The answer provides objective and unbiased information, for example, addresses all considerations of an issue, judgement does not appear to be swayed by considerations of self-interest or prejudice
Readability	The answer is easily readable, for example, organized, simple language, explanation of medical terms, and shorter sentences and paragraphs
Source credibility	The source of information is authoritative, for example, capable of being verified, does not seem to have commercial intent or personal agenda. Not applicable when no source is provided

As a check of the inter-rater reliability between 2 assessors, they independently rated the information quality of a common 10% (90 Q&As) of the data. We report inter-rater agreement measures for variables in the form that was used for model testing (ie, average indices) rather than the raw form [48,49]. We analyzed intraclass correlation coefficient (ICC) values using a single measurement, absolute agreement, 2-way random effects model. The ICC values for question quality index was 0.721 and for answer quality index was 0.764 with 95% confidence interval. Both quality indices exceeded 0.7, which indicates medium-to-good reliability [50].

Independent Variables

The independent variables are question quality and the platform design features.

To evaluate question quality, we used the criteria proposed by Harper et al [51] and Hsieh et al [26]: importance, perceived urgency, difficulty, question archival value, and writing quality. See Table 3 for definitions. PCA reveals only 1 eigenvalue >1.0 extracted from these measures of question quality (see Multimedia Appendix 3). Therefore, we built a single question quality composite measure by averaging the sum of the individual criteria ratings.

Each of the 9 selected Q&A platforms was reviewed by the first author to identify and record the presence of platform design features previously identified: incentives, Web 2.0 mechanisms, revenue models, expertise of participants, and question quality (see Table 4). LASSO allows high flexibility for inclusion of design features, as it does not impose constraints on the number of independent variables.

Table 3. Health question quality measures and definitions.

Quality criteria	Explanation
Importance	How seriously/sincerely did the questioner want an answer to the question? (eg, absence of reason for posting other than seeking information, eg, self-promotion/advertising product)
Perceived urgency	How urgently did the questioner want an answer to the question?
Difficulty	How difficult is the question to answer? (low and very low—anybody can answer the question; neither high nor low—an average high school–educated person is able to answer the question; high—someone with general medical background can answer the question; very high—specialist can answer the question)
Question archival value	Answer to this question will provide appropriate and adequate coverage of the issue to provide information of lasting/archival value to others
Writing quality	The question is well written (clear question, focused, and summarizes the issue)

Table 4. Variable descriptions.

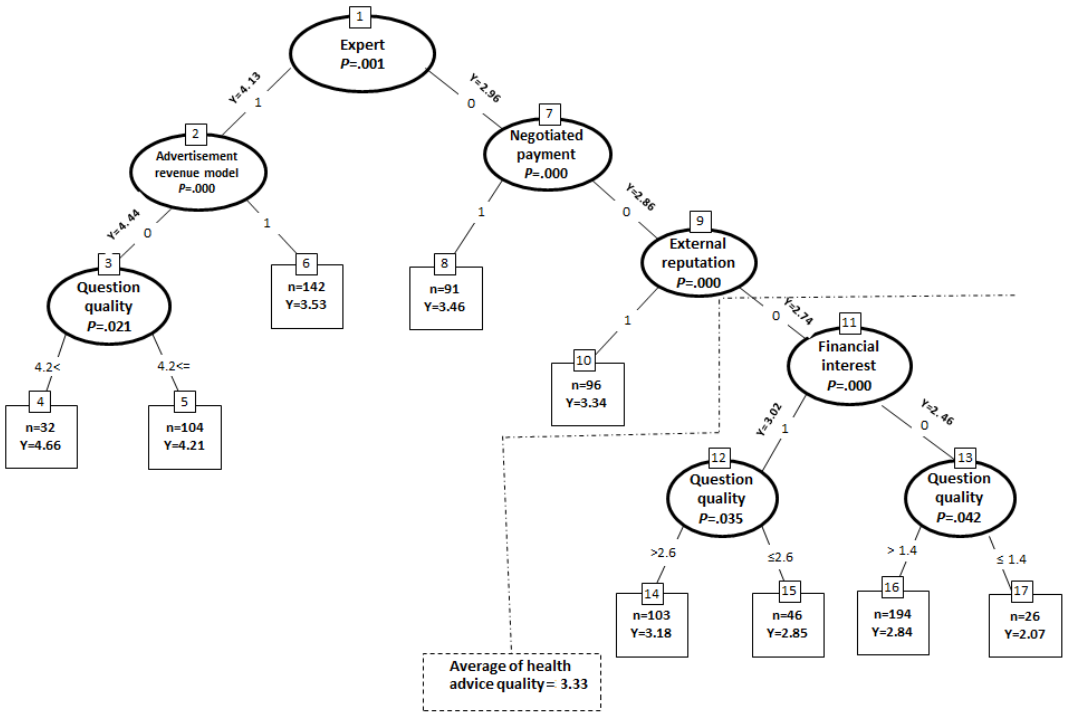
Category and variables	Explanations
Firm level	
Advertisement-based revenue; transaction-based revenue	Platform can use advertising or transaction-based model, both models, or neither model
Platform level	
Financial incentives	
Financial interest	Whether the information provider has any type of financial interest, either actually getting paid or with prospect of getting paid in future. In some cases, the advice provider was not actually paid but had the prospect of being hired by the platforms in the future
Payment for answering	Whether or not the information provider has been paid?
Variable payment scheme	The financial incentive could be paid in fixed or variable rate determined between the advice provider and asker
Amount of payment	How much money has been paid to information provider?
Nonfinancial incentives	
Virtual incentives	Whether any type of nonfinancial incentive, such as virtual points and credits, was involved?
Internal reputation system	Internal reputation system maintains and publicizes users' activity within a platform and their profiles
External reputation	External reputation is the identity and reputation of the users outside the platform in Web- and non-Web-based worlds
Mechanisms	
Web 2.0 mechanisms (voting, following)	These mechanisms reflect the feedback of users on each other's activity
User level	
Expert	The expertise of participants refers to their medical certification or their researching skill that is certified by the platform
Question quality	The quality of raised question

Analytical Methods

Our sample formed a wide dataset in that it included a large number of predictors with a comparatively small number of observations on the predictors. This feature of the sample makes it practically impossible to apply popular ordinary least square (OLS) regression techniques to model the data. A limitation of OLS regression for wide data concerns overfitting bias. Applying OLS to wide data can give rise to higher variance and poor out-of-sample predictions. LASSO provides suitable alternatives for modeling wide data [52]. It reveals and calculates the coefficients of predictively significant variables that minimize out-of-sample prediction error [53].

To provide a more nuanced explanation by identifying the interactions among predictors, we created an unbiased regression tree to study interactions among predictors (see Figure 1). The technique involves segmenting the predictor space into several regions. To make a prediction for a given observation, the method typically uses the mean of the training observations in the region to which it belongs. The splitting rules used to segment the predictor space are summarized in a tree structure. The advantage of regression tree is that it reveals conditional relationships, that is, how multiple combinations of design features are related to health advice quality, as it does not fit a linear model to the entire response space, it identifies conditional relationships between predictors and the response, and it presents the results in the form of a tree. Predictors appearing in higher layers of the tree or multiple times are predictively more significant than variables occurring in lower layers. Given the initial set of variables entering the tree, the variables absent from the tree do not improve prediction accuracy [54] (for analytical methods see Multimedia Appendix 4). To the best of our knowledge, this is one of very few studies that exploit LASSO and unbiased regression tree to determine design features that affect health advice quality.

Figure 1. The unbiased regression tree (N=834). "n" is the number of records; "Y" is the value of health advice quality.



Results

Least Absolute Shrinkage and Selection Operator

The composite measure of question answer quality had a mean of 3.33, standard deviation of 1.24, and median of 3.6 (N=834). The question quality composite measure had a mean of 3.28, standard deviation of 0.81, and median of 3.6 (N=900). We found that variables representing expertise of information providers, payment for answering questions, external reputation, revenue model, and question quality were the best predictors of health advice quality. The variable representing the use of an advertisement-based revenue model had negative association with health advice quality. Variables measuring nonfinancial incentives, internal reputation systems, and Web 2.0 mechanisms were not significant predictors. Table 5 presents LASSO coefficients, standard errors, and *P* values.

The initial analysis using LASSO explains the primary linear relationship between the best predictors and the outcome variable, health advice quality. The R-squared of the model was equal to 26%, which confirmed that the design of a Q&A platform could explain part of the advice quality. This is the basis for an overall understanding of the relationships between design features and health advice quality. However, at a more nuanced level of understanding, there may be interactive relationships between the predictors and the outcome that are apparent under certain conditions that can be explored using nonparametric techniques, such as regression tree, to extract the conditional relationships between design features and health advice quality.

Table 5. Least absolute shrinkage and selection operator results for the dependent variables.

Category and dependent variable	Regularization parameter	LASSO[a] coefficients	SE	*P* value
Intercept	2.24	—[b]	—	—
Firm level				
Advertisement-based revenue	−0.27	−0.25	0.07	<.001
Transaction-based revenue	0.11	0.26	0.09	.01
Platform level				
Financial incentives		0.40	0.08	<.001
Financial interest	0.00			
Payment for answering	0.25			
Negotiated payment	0.00			
Amount of payment	0.00			
Nonfinancial incentives		—	—	—
Virtual incentives	0.00			
Reputation		0.29	0.09	<.001
Internal	0.00			
External	0.10			
Web 2.0 mechanisms (voting, following)	0.00	—	—	—
User level				
Certify	0.00	—	—	—
Expert	0.59	0.48	0.07	<.001
Question quality	0.06	0.12	0.03	<.001
Number of records	834	—	—	—
Root-mean-square error	0.82	—	—	—
R-squared	0.26	—	—	—

[a]LASSO: least absolute shrinkage and selection operator.

[b]Not applicable.

Unbiased Regression Tree

The result of the regression tree is summarized in Figure 1 and shows that expertise was the most significant factor in predicting health advice quality because it appeared at the root node, supporting the LASSO results. In presence of experts, the second layer (node 2) distinguished between platforms where experts responded to health queries and advertising played no role and those where experts responded but advertising was a part of the revenue model. In the former case, there was further differentiation by question quality: when the mean question quality exceeded 4.2 (node 4), the mean answer quality was the highest (4.66). Question quality ≤4.2 (node 5) was associated with lower-quality health advices. Where advertising plays a role (node 6), answer quality is lower, that is, lower-quality answers were associated with platforms where no financial incentive was offered and providing information was on voluntary basis for philanthropic purposes of helping others or building external reputation. The interesting point is that, even so, the quality rating was above average for the dataset. So, although the advertisement-based revenue model had a negative relationship with quality, the strong effect of experts could alleviate this effect, and above-average quality was produced.

The other side of the tree concerns the absence of experts to provide answers. In the absence of experts, whether the advice provider receives any financial rewards and the platform allows the asker and answerer to directly agree on a price rather than establishing a fixed price per question (node 7) appeared as the second most important factor, and advice quality was still above average for the dataset (node 8). In node 9 where experts and variable payment scheme were absent, external reputation was associated with higher than average advice quality. In the next level where external reputation was absent (node 11), financial interest showed up. It means that advice providers with the prospect of getting paid in the future generated higher quality health advices compared with those who have no financial interest. Here, higher-quality questions were associated with higher-quality heath advices (node 14), although the quality is now below average for the dataset. The lowest quality health advice was generated in the absence of experts, motivation of external reputation, and any type of financial interest where low-quality information was raised (node 17).

Discussion

Overview

This study investigates the complexity of the relationship between design features of Web-based platforms and the quality of health advice generated in them. Theoretical and empirical evidence suggest a wide range of potential design features at different levels associated with health advice quality. The LASSO selects the design features that best predict the quality, and the regression tree identifies interactions among the design features.

Firm-Level Features

Revenue Model

Higher health advice quality is associated with the transaction-based revenue model, whereas lower advice quality is associated with an advertisement-based revenue model. This finding reinforces the findings of Harper et al [22], suggesting answer quality is superior in fee-based Web-based Q&A platforms than free sites. Taken on its own, this has worrying implications for these platforms to be instrumental in equality of Web access to quality health care.

Notwithstanding this intuitive result, the results also suggest that external reputation, expertise of the answerer, and question quality are also associated with higher quality. The revenue model is negatively associated in the LASSO regression models, which is in line with the worries regarding the advertisement-based revenue model, that is, these platforms tend to maximize the number of visitors at the expense of advice quality. On such platforms, no restrictions are placed on question quality or subject of postings; neither are users' efforts for creating high-quality advice recognized nor compensated. Although this is a common issue in all types of platforms, the potential consequences of poor advice, for example, incorrect diagnosis, users employing incorrect treatments, or ignoring correct treatments, are more serious in the health domain.

Nonetheless, as suggested in the introduction, the situation is more nuanced. The regression tree analysis refines our view about the advertisement-based model; we find indications of answer quality on these platforms that exceeds the average when experts give answers. In other words, the effect of experts in the health domain is so strong that the negative effect of the advertisement-based revenue model is controlled.

Platform Level

The results reveal that financial incentives are very influential in forming quality of health advice in different circumstances. First, there is a linear relationship between *payment for answering* and quality of health advice. Second, paid experts are providing higher quality than unpaid experts. Third, a variable payment scheme is associated with above-average quality advice. Finally, lay users with the prospect of getting paid in the feature are providing higher-quality answers comparing with those who have no financial interest at all.

Kissick [55] argues at a general level that quality, cost, and access are 3 essential, but competing, features of health care systems. This study confirms this argument in accessing Web-based health advice. High-quality advice is primarily accessible for payment; thus, access is compromised.

Web-based platforms such as eBay, TripAdvisor, and Amazon ask current users of their products/services to evaluate the quality of them after consumption. This feedback is aggregated and enables more informed decisions of future users. However, the usefulness of such mechanisms is questionable in health domain because of the nature of health advice as a *credence good*. Users can evaluate the quality of experience goods such as clothing, furniture, restaurant, hotel, etc. However, the quality of credence goods such as health care advice is difficult to evaluate for lay users even after consumption because they lack a medical background [56]. This undermines the effectiveness of any mechanism that works based on users' feedback in the health domain. Our results confirm that there is little association between these mechanisms, for example, Web 2.0 mechanisms and reputation systems and an external, independent assessment of health advice quality.

On the Web-based health platforms that rely purely on such mechanisms, this issue can result in the situation where poor-quality crowds out higher quality [57]. This means that in absence of a system that signals quality, contributing high-quality advice is not recognized. This leaves no motivation for advice providers to share quality advice, resulting in the platform comprising low-quality contributions.

Similarly, there is no evidence for effectiveness of virtual incentives. This supports the theory indicating that repeated positive rewards lose their value over time [58]. Therefore, reputation management may not need to be highly emphasized on Web-based information platforms because it may engage people at the early stage of being a platform user but be of value for long-term users, especially advanced answerers [28,38].

However, the results strongly suggest that above-average health advice quality by nonexperts can be produced when external reputation is used on the platform.

User Level

Expertise of Respondents

Crowd sourcing platforms rely on small contributions by a large number of people being superior to the contributions of a few experts [59]. However, our analyses suggest that in health care, advice provider expertise is the most effective predictor of answer quality. Experts, whether incentivized by financial or social rewards, are associated with the highest-quality answers.

It is surprising to see that volunteer health experts provide lower-quality health advice comparing with those who are paid. This might be explained by the lack of face-to-face interaction on Web-based platforms, and thus, lack of feedback on helpfulness of the contribution to receivers that decreases the satisfaction they get for helping others.

Question Quality

Altruism and enjoyment are among the top-ranked motivations of participation [34,38]. High question quality means clearer questions. Therefore, information providers have higher motivation to provide higher-quality answers to high-quality

questions. Our analyses support this view as high-quality answers are associated with high-quality questions, consistent with Hsieh et al [35] and Harper et al [51].

Across Levels

Regression tree analysis results show that a combination of experts, financial incentives, and high-quality questions predicts the highest health advice quality, suggesting that the highest quality is produced when experts are motivated by a combination of personal gain in the form of financial incentives, high enjoyment, and altruism provoked by high-quality questions. Recruiting unpaid experts, incentivizing lay users financially while using variable payment schemes, and incorporating external reputation are all associated with higher than average health care advice quality. Lowest quality advice occurs where there are no experts, any form of financial incentive is absent, and question quality is poor.

The tree analysis provides further insights. First, even paid experts provide higher-quality answers to higher-quality queries. Second, if an advertisement-based revenue platform can recruit expert respondents, this is related to answer quality beyond the average. In the absence of financial incentives and experts, external reputation is the best predictor of advice quality. Finally, the question quality has an effect, even on platforms where the clusters of features associated with poor quality are present.

Implications

The main contribution of this study is providing insights on the multiple relationships between the platform design and health care advice quality, and thus on Web-based health information platform design. Therefore, it has implications for health policy makers who attempt to address quality concerns and those concerned with access to Web-based health care information as a means for doing more for less for more people.

An important policy concern is inequality of access to quality advice on the internet. The findings show that the highest-quality advice is accessible for payment; it is not available for free, although only 2% of those seeking health information pay for it [60]. The connection between question quality and answer quality increases the danger of a triple hazard for low-income lower-literacy groups because of their lower ability to produce a quality question combined with a lower understanding of what is likely to be a lower-quality answer. Groups with low income and/or lower literacy are disadvantaged in accessing high-quality advice, increasing risks of electronic health (eHealth) inequality.

However, there is a potential to provide the highest-quality health advice for free, if the platform pays the recruited experts but does not charge the consumers. YouTube provides free access to all its videos but pays a fraction of its revenue from advertisements to video creators. Similarly, if the health platform pays experts from alternative revenue sources such as advertisement, selling data, and government subsidization, the highest quality can be generated and equally accessible for free.

In addition, designers can also consider the needs of the less well-educated, with low literacy or language barriers, for example, by providing voice or translation facilities, as Chu et al [61] recommended for other eHealth platforms.

The findings cast doubts on the effectiveness of common *internal* quality assessment features, such as voting and reputation systems, in the health domain and suggest using external reputation of the advice providers outside the platform.

Despite evidence that users form their own perceptions of information quality, platforms with objectively low advice quality are still available and viable, suggesting they answer some user needs, perhaps less available elsewhere. Lower question/answer quality is associated with less restrictions on asking or answering questions, these may go *off topic* and provide social support from fellow sufferers not found in more formal platforms, and the sheer volume of users may assist in assuaging feelings of facing an illness alone. Huxley et al [62] also report that marginalized groups find *stigmatizing reactions* a barrier. This may occur in more formal platforms, however, inadvertently. Studies of eHealth inequalities find that low literacy, low education, and other language difficulties are barriers to health information use [63]. Where answers are not provided by experts, the language may be less formal and more easily understood.

Limitations and Future Research

Our results show 26% explained variance. However, in a study of this kind with noisy, high-variability data where R-squared values are likely to be low, we found the explained variance for the emergent features was significantly different from 0, indicating a statistically significant explanatory power. Hence, we believe our results show significant patterns that provide valuable information of practical significance from which important conclusions can be drawn. Nonetheless, our sample is limited in the sense that it only captures, perhaps partially, design features of the existing platforms. There are a host of other factors that are likely to affect information quality. An example is the platform popularity. It might be that higher-quality information providers are more likely to join popular platforms. The reputation of the platform owner such as Yahoo! might also impact the decision to join a platform. Although we did not have access to such data, including such factors might further explain the variation in quality. The text also mentions the case of donation- or charity-based Q&A platforms. Recent advances in text recognition have also paved the way for effective information aggregation and comparing the quality of an advice with past answers to similar questions. Although we faced challenges in including such factors, they are likely to add to the explanatory power of the model. The choice of design features was also limited to existing platforms. There might be alternative designs that better facilitate generation of high-quality information. An example is given in the text: donation- or charity-based revenue models might lead to higher-quality answers and extend access. In general, as in other platforms, an important avenue for learning about better design is experimentation. Future research should conjecture new designs and conduct A/B testing to learn about the effectiveness of the design features. Major platforms such as Amazon, Alphabet, eBay, and Facebook have turned to experimentation to improve their platforms. This can equally

benefit the design of health information platforms. Advances in artificial intelligence and text mining techniques are bound to impact health information platforms. These advances will raise challenges for future research. Methodologically, future research can use recently developed causal regression trees to identify design features that drive information quality. This research rated health advice quality from medical point of view; future research should investigate the user perspective concerning health advice quality, especially in view of potential barriers for disadvantaged or marginalized groups. Further research is needed to explain why volunteer unpaid health experts and paid experts provide uneven health quality advice in the Web-based world. Furthermore, we considered only English language platforms, leaving scope for future cross-cultural investigation. We used a robust search strategy to identify the design features; however, the inclusion of design feature was based on judgement of one coder, and in addition, we did not adopt *systematic search*.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Platforms.
[DOCX File , 42 KB - jmir_v22i1e13534_app1.docx]

Multimedia Appendix 2
Rating process.
[DOCX File , 113 KB - jmir_v22i1e13534_app2.docx]

Multimedia Appendix 3
Principal component analysis.
[DOCX File , 20 KB - jmir_v22i1e13534_app3.docx]

Multimedia Appendix 4
Analytical methods.
[DOCX File , 49 KB - jmir_v22i1e13534_app4.docx]

References

1. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
2. Buchanan H, Coulson NS. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269. [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]
3. Henderson EM, Keogh E, Eccleston C. Why go online when you have pain? A qualitative analysis of teenagers' use of the internet for pain management advice. Child Care Health Dev 2014 Jul;40(4):572-579. [doi: 10.1111/cch.12072] [Medline: 23663232]
4. Skinner H, Biscope S, Poland B. Quality of internet access: barrier behind internet use statistics. Soc Sci Med 2003 Sep;57(5):875-880. [doi: 10.1016/s0277-9536(02)00455-0] [Medline: 12850112]
5. von Rosen AJ, von Rosen FT, Tinnemann P, Müller-Riemenschneider F. Sexual health and the internet: cross-sectional study of online preferences among adolescents. J Med Internet Res 2017 Nov 8;19(11):e379 [FREE Full text] [doi: 10.2196/jmir.7068] [Medline: 29117927]
6. Bhargava B. Welcome to BMJ Innovations. BMJ Innov 2014 Dec 10;1(1):1-2. [doi: 10.1136/bmjinnov-2014-000028]
7. Bhatti YA, Prime M, Harris M, Wadge H, McQueen J, Patel H, et al. The search for the holy grail: frugal innovation in healthcare from low-income or middle-income countries for reverse innovation to developed countries. BMJ Innov 2017 Nov 4;3(4):212-220. [doi: 10.1136/bmjinnov-2016-000186]
8. Kitchens B, Harle CA, Li S. Quality of health-related online search results. Decis Support Syst 2014 Jan;57:454-462. [doi: 10.1016/j.dss.2012.10.050]
9. Corritore C, Wiedenbeck S, Kracher B, Marble R. Online trust and health information websites. Int J Hum Comput Interact 2012;8(4):92-115. [doi: 10.4018/jthi.2012100106]
10. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 7;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
11. Pletneva N, Vargas A, Boyer C. How do general public search online health information. Health Net Found 2011:2-38 [FREE Full text]

12. Lupiáñez-Villanueva F. Health and the internet: beyond the quality of information. Rev Esp Cardiol 2011 Oct;64(10):849-850 [FREE Full text] [doi: 10.1016/j.recesp.2011.06.014] [Medline: 21839571]
13. Ferguson T. From patients to end users. Br Med J 2002 Mar 9;324(7337):555-556 [FREE Full text] [doi: 10.1136/bmj.324.7337.555] [Medline: 11884301]
14. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
15. Deshpande A, Jadad AR. Trying to measure the quality of health information on the internet: is it time to move on? J Rheumatol 2009 Jan;36(1):1-3. [doi: 10.3899/jrheum.081101] [Medline: 19208527]
16. Kanthawala S, Vermeesch A, Given B, Huh J. Answers to health questions: internet search results versus online health community responses. J Med Internet Res 2016 Apr 28;18(4):e95 [FREE Full text] [doi: 10.2196/jmir.5369] [Medline: 27125622]
17. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
18. Fiedler M, Sarstedt M. Influence of community design on user behaviors in online communities. J Bus Res 2014 Nov;67(11):2258-2268. [doi: 10.1016/j.jbusres.2014.06.014]
19. Cabral L. Reputation on the internet. In: Peitz M, Waldfogel J, editors. The Oxford Handbook of the Digital Economy. Oxford, UK: Oxford University Press; 2012:343-335.
20. 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]
21. Bryman A, Bell E. Business Research Methods. Oxford, UK: Oxford University Press; 2015.
22. Harper FM, Raban D, Rafaeli S, Konstan JA, Carmel M. Predictors of Answer Quality in Online Q&A Sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI'08; April 05-10, 2008; Florence, Italy p. 865-874. [doi: 10.1145/1357054.1357191]
23. Enders A, Hungenberg H, Denker HP, Mauch S. The long tail of social networking. Eur Manag J 2008 Jun;26(3):199-211. [doi: 10.1016/j.emj.2008.02.002]
24. Chen Y, Ho TH, Kim YM. Knowledge market design: a field experiment at Google Answers. J Public Econ Theory 2010;12(4):641-664. [doi: 10.1111/j.1467-9779.2010.01468.x]
25. Wang J, Ghose A, Ipeirotis P. Bonus, Disclosure, and Choice: What Motivates the Creation of High-Quality Paid Reviews? In: Proceedings for the Thirty Third International Conference on Information Systems. 2012 Presented at: ICIS'12; December 16-19, 2012; Orlando.
26. Hsieh G, Kraut RE, Hudson SE. Why Pay? Exploring How Financial Incentives Are Used for Question & Answer. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems. 2010 Presented at: CHI'10; April 10-15, 2010; Atlanta, Georgia, USA. [doi: 10.1145/1753326.1753373]
27. Chang H, Chuang S. Social capital and individual motivations on knowledge sharing: participant involvement as a moderator. Inf Manag 2011 Jan;48(1):9-18. [doi: 10.1016/j.im.2010.11.001]
28. Khansa L, Ma X, Liginlal D, Kim S. Understanding members' active participation in online question-and-answer communities: a theory and empirical analysis. J Manag Inform Syst 2015 Aug 28;32(2):162-203. [doi: 10.1080/07421222.2015.1063293]
29. Hung SY, Durcikova A, Lai HM, Lin WM. The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. Int J Hum-comput St 2011 Jun;69(6):415-427. [doi: 10.1016/j.ijhcs.2011.02.004]
30. Tausczik YR, Pennebaker JW. Predicting the Perceived Quality of Online Mathematics Contributions From Users' Reputations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 07-12, 2011; Vancouver, BC, Canada p. 1885-1888. [doi: 10.1145/1978942.1979215]
31. Reavley NJ, Mackinnon AJ, Morgan AJ, Alvarez-Jimenez M, Hetrick SE, Killackey E, et al. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychol Med 2012 Aug;42(8):1753-1762. [doi: 10.1017/S003329171100287X] [Medline: 22166182]
32. Giles J. Internet encyclopaedias go head to head. Nature 2005 Dec 14;438(7070):900-901. [doi: 10.1038/438900a]
33. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]
34. Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci 2011 Nov 1;63(3):543-557. [doi: 10.1002/asi.21676]
35. Hsieh G, Counts S. Mimir: A Market-based Real-time Question and Answer Service. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA, USA p. 769-778. [doi: 10.1145/1518701.1518820]
36. Ludwig S, de Ruyter K, Mahr D, Wetzels M, Brüggen E, de Ruyck T. Take their word for it: the symbolic role of linguistic style matches in user communities. MIS Quarterly 2014 Apr 4;38(4):1201-1217. [doi: 10.25300/misq/2014/38.4.12]
37. Walther JB, Wang Z, Loh T. The effect of top-level domains and advertisements on health web-site credibility. J Med Internet Res 2004 Sep 3;6(3):e24 [FREE Full text] [doi: 10.2196/jmir.6.3.e24] [Medline: 15471750]

38. Oh S, Syn SY. Motivations for sharing information and social support in social media: a comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. J Assn Inf Sci Tec 2015 Apr 21;66(10):2045-2060. [doi: 10.1002/asi.23320]
39. Anderka M, Stein B, Lipka N. Predicting Quality Flaws in User-generated Content: The Case of Wikipedia. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2012 Presented at: SIGIR'12; August 12-16, 2012; Portland, Oregon, USA p. 981-990. [doi: 10.1145/2348283.2348413]
40. Shah C, Kitzie V, Choi E. Questioning the Question -- Addressing the Answerability of Questions in Community Question-Answering. In: Proceedings of the 2014 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS'14; January 6-9, 2014; Hawaii p. 1386-1395. [doi: 10.1109/hicss.2014.180]
41. Google. URL: https://www.google.com/ [accessed 2014-01-01]
42. Yahoo. URL: https://www.yahoo.com/ [accessed 2014-01-01]
43. Bing. URL: https://www.bing.com/ [accessed 2014-01-01]
44. Ask. URL: https://www.ask.com [accessed 2014-01-01]
45. Kitzie V, Choi E, Shah C. Analyzing Question Quality Through Intersubjectivity: World Views and Objective Assessments of Questions on Social Question-answering. In: Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries. 2013 Presented at: ASIST'13; November 1-5, 2013; Montreal, Quebec, Canada. [doi: 10.1002/meet.14505001052]
46. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
47. Oh S, Yi YJ, Worrall A. Quality of Health Answers in Social Q&A. In: Proceedings of the American Society for Information Science and Technology. 2013 Presented at: ASIS&T'12; 2013; Baltimore, MD. [doi: 10.1002/meet.14504901075]
48. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. New York, NY: Springer; 2013.
49. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
50. 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]
51. Harper F, Moy D, Konstan J. Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA, USA p. 759-768. [doi: 10.1145/1518701.1518819]
52. Belloni A, Chernozhukov V, Hansen C. Inference on treatment effects after selection among high-dimensional controls. Rev Econ Stud 2013 Nov 24;81(2):608-650. [doi: 10.1093/restud/rdt044]
53. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol 2018 Dec 5;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
54. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 2006 Sep;15(3):651-674. [doi: 10.1198/106186006x133933]
55. Kissick WL. Medicine's Dilemmas: Infinite Needs Versus Finite Resources. New Haven, CN: Yale University Press; 1994.
56. Darby MR, Karni E. Free competition and the optimal amount of fraud. J Law Econ 1973 Apr;16(1):67-88. [doi: 10.1086/466756]
57. Akerlof GA. The market for 'lemons': quality uncertainty and the market mechanism. Q J Econ 1970 Aug;84(3):488. [doi: 10.2307/1879431]
58. Homans G. Social Behavior: Its Elementary Forms. London: Routledge & Kegan Paul; 1961.
59. Surowiecki J. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. New York, NY: Doubleday; 2004.
60. Fox S, Duggan M. Pew Research Center. Health Online 2013 URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2018-07-22]
61. Chu A, Huber J, Mastel-Smith B, Cesario S. 'Partnering with Seniors for Better Health': computer use and internet health information retrieval among older adults in a low socioeconomic community. J Med Libr Assoc 2009 Jan;97(1):12-20 [FREE Full text] [doi: 10.3163/1536-5050.97.1.003] [Medline: 19159002]
62. Huxley CJ, Atherton H, Watkins J, Griffiths F. Digital communication between clinician and patient and the impact on marginalised groups: a realist review in general practice. Br J Gen Pract 2015 Dec;65(641):e813-e821 [FREE Full text] [doi: 10.3399/bjgp15X687853] [Medline: 26622034]
63. Feng Y, Xie W. Digital divide 2.0: the role of social networking sites in seeking health information online from a longitudinal perspective. J Health Commun 2015;20(1):60-68. [doi: 10.1080/10810730.2014.906522] [Medline: 25119019]

Abbreviations

eHealth: electronic health
ICC: intraclass correlation coefficient

LASSO: least absolute shrinkage and selection operator
OLS: ordinary least square
PCA: principal component analysis
Q&A: question and answer

Edited by G Eysenbach; submitted 29.01.19; peer-reviewed by E Das, J Gauly, R Haase, W Pian, A Majmundar; comments to author 28.03.19; revised version received 26.06.19; accepted 02.08.19; published 15.01.20.

Please cite as:
Ameri F, Keeling K, Salehnejad R
You Get What You Pay for on Health Care Question and Answer Platforms: Nonparticipant Observational Study
J Med Internet Res 2020;22(1):e13534
URL: https://www.jmir.org/2020/1/e13534
doi:10.2196/13534
PMID:31939741

Review

Tools to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals: Systematic Review

Gerlinde Lenaerts[1], MSc, PhD; Geertruida E Bekkering[1], MSc, PhD; Martine Goossens[1], MPhil; Leen De Coninck[1,2,3], MSc; Nicolas Delvaux[2], MD; Sam Cordyn[4], MSc; Jef Adriaenssens[5], MSc, PhD; Patrick Vankrunkelsven[1,2], MD, PhD

[1]Belgian Centre for Evidence Based Medicine (CEBAM), Leuven, Belgium

[2]Department of Public Health and Primary Care, Katholieke Universiteiti Leuven, Leuven, Belgium

[3]Artevelde Hogeschool, Ghent University Association, Ghent, Belgium

[4]Federation of the White and Yellow Cross of Flanders, Brussels, Belgium

[5]Belgian Health Care Knowledge Centre, Brussels, Belgium

Corresponding Author:
Gerlinde Lenaerts, MSc, PhD
Belgian Centre for Evidence Based Medicine (CEBAM)
Kapucijnenvoer 33
blok J bus 7001
Leuven, 3000
Belgium
Phone: 32 16377273
Email: gerlinde.lenaerts@cebam.be

Abstract

Background: User-friendly information at the point of care should be well structured, rapidly accessible, and comprehensive. Also, this information should be trustworthy, as it will be used by health care practitioners to practice evidence-based medicine. Therefore, a standard, validated tool to evaluate the trustworthiness of such point-of-care information resources is needed.

Objective: This systematic review sought to search for tools to assess the trustworthiness of point-of-care resources and to describe and analyze the content of these tools.

Methods: A systematic search was performed on three sources: (1) we searched online for initiatives that worked off of the trustworthiness of medical information; (2) we searched Medline (PubMed) until June 2019 for relevant literature; and (3) we scanned reference lists and lists of citing papers via Web of Science for each retrieved paper. We included all studies, reports, websites, or methodologies that reported on tools that assessed the trustworthiness of medical information for professionals. From the selected studies, we extracted information on the general characteristics of the tools. As no standard, risk-of-bias assessment instruments are available for these types of studies, we described how each tool was developed, including any assessments on reliability and validity. We analyzed the criteria used in the different tools and divided them into five categories: (1) author-related information; (2) evidence-based methodology; (3) website quality; (4) website design and usability; and (5) website interactivity. The percentage of tools in compliance with these categories and the different criteria were calculated.

Results: Included in this review was a total of 17 tools, all published between 1997 and 2018. The tools were developed for different purposes, from a general quality assessment of medical information to very detailed analyses, all specifically for point-of-care resources. However, the development process of the tools was poorly described. Overall, seven tools had a scoring system implemented, two were assessed for reliability only, and two other tools were assessed for both validity and reliability. The content analysis showed that all the tools assessed criteria related to an evidence-based methodology: 82% of the tools assessed author-related information, 71% assessed criteria related to website quality, 71% assessed criteria related to website design and usability, and 47% of the tools assessed criteria related to website interactivity. There was significant variability in criteria used, as some were very detailed while others were more broadly defined.

Conclusions: The 17 included tools encompass a variety of items important for the assessment of the trustworthiness of point-of-care information. Overall, two tools were assessed for both reliability and validity, but they lacked some essential criteria

for the assessment of the trustworthiness of medical information for use at the point-of-care. Currently, a standard, validated tool does not exist. The results of this review may contribute to the development of such an instrument, which may enhance the quality of point-of-care information in the long term.

Trial Registration: PROSPERO CRD42019122565; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=122565

(J Med Internet Res 2020;22(1):e15415) doi:10.2196/15415

KEYWORDS

evidence-based medicine; evidence-based practice; point-of-care systems; health care quality; internet information; information science; systematic review

Introduction

Evidence-based medicine is one of the cornerstones of high-quality health care. This conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients [1] should be facilitated to ensure effective and efficient patient care. With a continuously increasing body of scientific evidence, it is not feasible for health care professionals to access and review the best evidence themselves regularly and independently. Furthermore, they have little time to process large quantities of information during their consultation with patients [2]. Therefore, health care professionals need good quality information that is also user-friendly. This type of information is labeled point-of-care information [3,4], and it is well-structured, rapidly accessible, and comprehensive information for use at the specific point in the workflow when health care professionals and patients interact [3].

Health care professionals routinely use clinical guidelines as reliable sources of information to support their clinical decision-making. Guidelines are statements that include recommendations that are intended to optimize patient care, and that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options [5]. This combination of an assessment of quality of evidence and the benefits and harms means guidelines are most suited to guide clinical decision-making. Furthermore, a validated instrument is available to assess the quality of guidelines [6]. This instrument, known as AGREE II (Appraisal of Guidelines for Research and Evaluation), was developed for the assessment of the validity and trustworthiness of clinical guidelines and is nowadays recognized as an international standard. The use of such an instrument enhances the quality of guidelines [7,8]; however, for many clinical problems or health care professions, there are no or limited guidelines available. In that case, one depends on other information sources. Thanks to the internet, a vast amount of information is accessible within a few mouse clicks, but identification of the most relevant information and assessment of its quality and transparency is indispensable when used in clinical practice. Although different instruments for assessment of the methodological quality of systematic reviews [9-11] or individual studies [12,13] do exist, these instruments are not appropriate for evaluation of the trustworthiness of point-of-care information. Banzi et al [3] reviewed online point-of-care information summary providers. They developed a tool to evaluate the breadth, content development, and editorial policy against their claims of being “evidence-based.” However, this tool was never tested on validity and reliability.

We aimed to search for a valid tool to assess the trustworthiness of point-of-care information. To this end, we performed a systematic review to identify existing tools and examined their validity and reliability.

Methods

Overview

We performed a systematic review using the standards for systematic reviewing reported by Cochrane [14], and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used for the reporting of our findings [15]. The protocol of this review was registered at PROSPERO (CRD42019122565).

Search Strategy

To identify tools, we used three sources of information. First, we searched the internet for institutes or initiatives that worked on the trustworthiness of health information. Second, we searched Medline (via PubMed) for relevant literature. A search from the inception of the database to June 2019 was conducted to identify the studies of interest. A search string was built using the concepts of trustworthiness and point-of-care information (Textbox 1). Terms within a concept were combined using the Boolean operator ‘OR.’ Then, the terms between the concepts were combined with the Boolean operator ‘AND.’ Lastly, we scanned the reference lists and lists of all citing papers via Web of Science for each retrieved paper, to identify additional tools that were not found in the previous searches.

Textbox 1. Concepts used to build the search string.

Concept 'trustworthiness.'

- Mesh-terms: methods; standards (subheading); healthcare evaluation mechanisms; evaluation studies as topic; health care quality, access, and evaluation; reproducibility of results
- Free-text words: methodological quality, quality standards, evidence-based methodology, editorial quality, evaluation, validity, reliability

Concept 'point-of-care.'

- Mesh-terms: health information systems; point-of-care systems; medical informatics, consumer health informatics;
- Free-text words: (web-based or electronic or online or internet) and health information; e-Health, e-Health information, point-of-care services, point-of-care information

Inclusion and Exclusion Criteria

We included all studies, reports, websites, or methodologies that reported on tools, including checklists and criteria, to assess the trustworthiness of medical information for health care professionals. We used the following criteria:

1. Tools had to evaluate point-of-care information or resources for professionals. The definition of point-of-care information was web-based medical compendia specifically designed to deliver predigested, rapidly accessible, comprehensive, periodically updated, and evidence-based information (and guidance) to clinicians [3]. We excluded tools to assess the quality of information for patients, as well as tools that assessed the quality of systematic reviews or other primary studies.
2. Tools had to evaluate trustworthiness. Trustworthiness represented features that made users trust the information, including methodological quality and editorial transparency. The tools that only assessed user-friendliness were excluded.
3. Tools had to be published by multiple authors or an organization.
4. Tools had to be freely available. In the case of websites that contained multiple tools, they were separated by their methodology.
5. Additionally, we excluded tools to assess the quality of mobile applications.

Selection of Articles and Web Pages

Tools were selected by two researchers (GB, GL) independently. The selection of journal articles was made in two steps: (1) all titles and abstracts were compared against the selection criteria; and (2) the full texts of potential eligible articles were retrieved and subsequently compared against the inclusion and exclusion criteria. The two researchers resolved discrepancies in selection by discussion and consensus.

Assessment of Methodological Quality

To date, there are no standards to assess the methodological quality of tools to determine the trustworthiness of point-of-care information. Therefore, we could not perform a standard risk-of-bias assessment on each tool. However, we checked each tool for potential risk of bias in the developmental phase, including looking for a lack of validity and performing a reliability assessment. Also, we extracted details on the development of the tools.

Data Extraction and Analysis

A data overview table was used to extract data from the available tools (Textbox 2). We noted the general characteristics and described the purpose for which a tool was developed and the criteria and scoring systems used. The data extraction was then performed by one researcher (GL) and checked by a second researcher (GB). Discrepancies were identified and resolved through discussion. Based on the data overview table, both the similarities and differences of the characteristics of the tools used to assess the trustworthiness of point-of-care information were analyzed. To examine possible overlap between tools, we listed all criteria and mapped them into general ones, then divided those into five main categories: (1) author-related information; (2) evidence-based methodology; (3) website quality; (4) website design and usability; and (5) website interactivity (see Multimedia Appendix 1). Based on these categories and criteria, we described the characteristics of the tools using descriptive statistics.

Textbox 2. Data overview table.

Characteristics of the tool

- Name
- Aim
- Developer
- Is there a sum score of final combined judgment?
- Description (items, elements)
- Description (scoring method)
- Remarks

Development of the tool

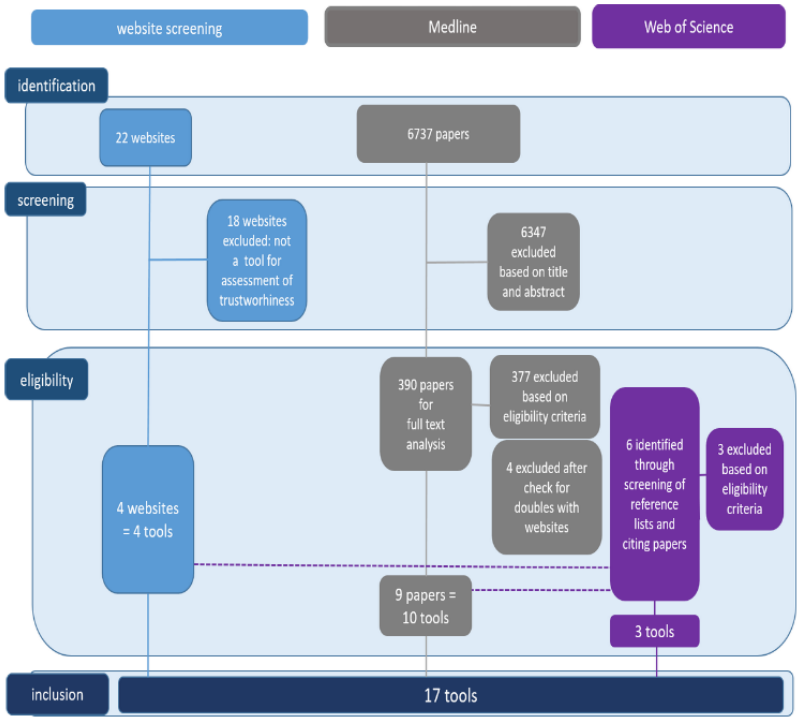
- How was the tool developed? (descriptive)
- For which purpose was the tool developed? (descriptive)
- Was the tool assessed for validity? (Y/N)
- Was the tool assessed for reliability? (Y/N)
- Other relevant details (descriptive)

Results

Search Strategy

The flow chart shown in Figure 1 summarizes the results of the systematic search. After identification of relevant websites and titles and screening of abstracts and full texts for eligibility, 16 papers that reported on tools or criteria to assess the trustworthiness of point-of-care information used by health care professionals were included. One study reported on more than one tool [16]. Finally, 17 tools were included in this review.

Figure 1. Search strategy.



General Characteristics of the Tools

Table 1 provides an overview of the general characteristics of the included tools. All tools were developed between 1997 and 2018, and they originated from the United States, Canada, Europe, Switzerland, Singapore, and Iran. The Health on the Net (HON) code [17] and the electronic health (eHealth) Code of Ethics [18] are both codes of conduct and the result of international collaboration. They were developed based on discussion and consensus with international expert panels and underwent peer review. The HONcode aims to control the quality of health information on the internet and provides a quality seal for HONcode-certified websites [17]. The eHealth Code of Ethics provides guiding principles to understand the

risk and potential of health information on the internet for professionals, producers, and consumers, and aims to contribute to high-quality information in this way [18].

The Silberg [19], Kapoun [20], Gillois [21], Jiang criteria [22], and CART (Completeness, Accuracy, Relevance, Timeliness) [23] tools aimed to critically appraise and evaluate the quality, credibility, and appropriateness of health information on the internet. The development process of these tools was poorly described. The authors relied on existing criteria for the quality assessment of the information [20-23], or the critical thinking process of the authors was the basis for the definition of the criteria [19]. Likewise, the Sandvik scale [24], QUEST (Quality Evaluation Scoring Tool) [25], and the 11-Point Quality Assessment Scale [26] were developed for the same purpose, but they have a scoring system implemented. The Aslani criteria [16] are based on the HONcode, Silberg criteria, Kapoun criteria, Sandvik scale, and the Health Information Technology Institute (HITI) criteria. However, these criteria were excluded from this review because they are not available anymore.

The AMA (American Medical Association) developed principles [27] as guidelines meant to apply to all AMA websites, but they were also intended to guide the creators of websites that provide medical information for professionals and consumers. The principles are developed and regularly reviewed by AMA staff members and an external advisory panel of experts.

The Grid ULiège [28,29] and the Trumble Tool [30] are the most comprehensive tools and were developed for the analysis and evaluation of medical websites. An Excel-based evaluation form allows the calculation of a final, weighted score based on 38 or 23 items, respectively. The Trumble tool was specifically developed to evaluate evidence-based medical tools at point-of-care. The Banzi tool [3] aims to evaluate and score the breadth, content development, and editorial policy for point-of-care summaries against their claims of being evidence-based. Finally, OncoRX-IQ [31] is a tool developed for the assessment of the quality of information of online drug databases for anticancer drug interactions. The Banzi tool [3], the Trumble tool [26], and the 11-Point Quality Assessment Scale [30] are the only tools that specify the evaluation of evidence-based principles of online health information. The definitions of the content of these tools were done by researchers who arbitrarily postulated criteria that, to their opinion, would best describe the quality of point-of-care information. Only 4/17 (24%) tools were assessed for reliability [21,25,31] and only 2/17 (12%) for validity [25,26] (Table 1).

Table 1. Characteristics of the tools.

Name of tool	Language	Date of publication	Country of origin	Number of items	Assessed for reliability or validity	Scoring system
Silberg criteria [19]	English	1997	United States	4	—	—
HONcode[a] [32]	English	1998	Based in Switzerland, international working group	8	—	—
Kapoun criteria [20]	English	1998	United States	5	—	—
Sandvik scale [24]	English	1999	Norway	7	—	Yes
Gillois criteria [21]	English	1999	France	9	Reliability	—
Joubert criteria [33]	English	1999	France	8	—	—
AMA[b] principles [27]	English	2000	United States	14	—	—
eHealth[c] Code of Ethics [18]	English	2000	United States, international working group (WHO[d]/PAHO[e])	17	—	—
Jiang criteria [22]	English	2000	United States	7	—	—
Grid ULiege [28,29]	English	2003	Belgium	38	—	Yes
CART[f] [23]	English	2006	United Kingdom	4	—	—
Trumble Tool [30]	English	2006	United Kingdom	23	—	yes
Banzi tool [3]	English	2010	Italy	10	—	yes
OncoRx-IQ [31]	English	2010	Singapore	19	Reliability	yes
11 Point Quality Assessment Scale [26]	English	2012	Canada	11	Reliability and validity	yes
Aslani criteria [16]	English	2014	Iran	10	—	—
QUEST[g] criteria [25]	English	2018	Canada	6	Reliability and validity	yes

[a]HON: health on the net.

[b]AMA: American Medical Association.

[c]eHealth: electronic health.

[d]WHO: World Health Organization.

[e]PAHO: Pan American Health Organization.

[f]CART: Completeness, Accuracy, Relevance, Timeliness.

[g]QUEST: Quality Evaluation Scoring Tool.

Content Analysis of the Tools

Multimedia Appendix 1 presents an overview of the 17 included tools with their criteria for the assessment of the trustworthiness of point-of-care information. Altogether, the tools cover 156 criteria. These were combined into 36 general criteria, mapped in five main categories: (1) author-related information with 4 related criteria; (2) evidence-based methodology with 15 related criteria; (3) website quality with 8 related criteria; (4) website design and usability with 7 related criteria; and (5) website interactivity. Some criteria described in the tools were broad and covered more than one general criterion and vice versa, whereas some criteria described in the tools were detailed and therefore summarized in one general criterion. For a few tools [3,18,27,28,30,31], we excluded some of the criteria because they were inappropriate for the assessment of trustworthiness or not applicable in the current context.

Multimedia Appendix 2 presents the prevalence of criteria in the 17 included tools. Overall, 14/17 tools (82%) of the tools addressed author-related information. Only the Joubert criteria, CART, and the 11-Point Quality Assessment Scale did not assess author-related information. All 17 tools (100%) addressed one or more items in the category of evidence-based methodology. The criteria "references to source data" (n=11; 65%) and "content is current and actual" (n=15; 88%) were the most frequently assessed in this category. A total of 12 tools (71%) assessed criteria related to website quality. The most frequently assessed criteria in this category were "transparent ownership" (n=9; 53%) and "financial information" (n=9; 53%). Website design and usability were evaluated by 12 tools (71%). The criterion "ease of use and navigation" (n=11; 65%) was the most frequently used. Website interactivity refers to functions that allow contact or discussion with the authors or site owners. This category was mentioned in 8 tools (47%).

Assessment of Reliability and Validity of Tools

The reliability and validity of the tools were scarcely reported. Interrater reliability was calculated by kappa coefficients [25,26], Kendall coefficients [31], or by calculation of a percentage of agreement between two researchers [21]. QUEST

was compared to three other criteria-related tools to calculate convergent validity. The quality scores generated by each pair of tools were compared to calculate Kendall tau-ranked correlation. The 11-Point Quality Assessment Scale stated that the tool was previously validated, but no information on the validation process could be found.

Discussion

Primary Findings

This review studied 16 articles that reported on 17 tools analyzing the trustworthiness of point-of-care information. Our main finding was that the trustworthiness of information is currently assessed and scored in different ways, as illustrated by essential differences in the number of criteria and the content addressed by the tools. This reveals the need for consistency and completeness in evaluating the quality of health information resources. Therefore, this review extends the current literature by giving an overview of existing tools, including their criteria and general characteristics.

To assess the trustworthiness of health care information, we need reliable tools that have been assessed on reliability and validity. Only QUEST [25] and the 11-Point Quality Assessment Scale [26] were assessed on both reliability and validity. However, QUEST only encompasses criteria on author-related information and evidence-based methodology and is therefore too concise for quality assessment of point-of-care information. The 11-Point Quality Assessment Scale was developed as a quality measure for online texts and covers criteria related to evidence-based methodology and usability. However, criteria for the assessment of author-related information and website quality, such as transparent ownership and financial disclosures, were missing, and the validation process was not described. The absence of reliable and validated tools is an essential finding of this review and a shortcoming in the field.

The criteria used in tools to assess the trustworthiness of medical information showed much variation. We encountered this variation when it became apparent that it would be difficult to structure all the original criteria and to reformulate general criteria. Some criteria overlapped with others, using slightly different terms, which illustrated the lack of uniformity and consistency in tools for quality assessment of point-of-care information.

Back in 2001, Risk and Dzenowagis [34] highlighted the complexity of health information on the internet and analyzed the major quality initiatives. A set of quality criteria for health information and credible enforcement tools were named as essential elements for successful quality programs. Nowadays, the need for a uniform, validated tool for the quality assessment of point-of-care resources is still present. Currently, the quality of most point-of-care information is low [3,35]. Risk has suggested tool-based evaluation of quality and third-party certification of compliance as critical mechanisms for quality improvement [34]. A valid tool may improve the quality of point-of-care information, as was also reported for guidelines [8].

Content of the Tools

All the tools have criteria to assess the evidence-based methodology used to summarize the information. Some use only two [16,19,27] while other tools have seven or more criteria in this category [3,18,26,28,29]. Perhaps the content of the items is more important than the number. For example, “reference to source data” and “content is current and actual” are frequently used, but often these criteria do not guarantee that an information source is truly evidence-based. Remarkably, only a few criteria fit the first three steps in evidence-based medicine: asking a good question, finding the best evidence, and appraising the evidence [36]. For example, “systematic reviews are preferred on primary studies,” “formal grading of evidence,” and “reporting of bias” are not standard and not addressed in the different tools.

A closer look at the weighted scoring system implemented in the Trumble tool [30] and the Grid ULiège [28] reveals that criteria related to the evidence-based methodology are the most important, as they receive the highest weight factor. The Trumble tools gives equal weight to criteria related to usability and currency, while Grid ULiège gives lower weight to criteria related to usability. Banzi et al [3] based their tool on criteria from research on systematic review reporting methods and peer-reviewed medical journals’ policies. The tool was developed to check point-of-care information against their claims of being evidence-based [3,4], which clarifies focus on this topic. The eHealth Code of Ethics [18], the Banzi tool [3], and the 11-Point Quality Assessment Scale [26] focus on the evidence-based aspects but have little or no attention for the category “website design and usability,” which is related to the point-of-care aspect of information. Health information sources that are difficult to navigate will likely be used less since time constraints are an important barrier for health care professionals [2,37]. Therefore, the ideal tool should find the right balance between evidence-based methodology and usability-related criteria.

The Grid ULiège include detailed criteria, but the descriptions were sometimes unclear and seemed to contain overlapping items. Conversely, other tools [17-20,24,25,27,31] addressed multiple content aspects in only one criterion. Some tools were very concise in terms of the number of criteria [19,23] and seemed insufficient for a thorough evaluation of medical information, while others were too extensive and detailed and were therefore difficult to use [28,29]. These findings show that an adequate definition of criteria, together with a rational number of criteria, is indispensable for the usability of a tool.

For a few included tools [3,18,27,28,30,31], some criteria were excluded because they were considered inappropriate for the assessment of trustworthiness or not applicable in the current context (eg, criteria related to electronic commerce and marketing or drug-specific criteria) (see Multimedia Appendix 2). The criterion “breadth and volume” was excluded because it would disadvantage information sources that were designed for one pathology or treatment. Moreover, the specificity of an information source did not necessarily affect the quality, and small volume sources may contain useful information for practitioners.

Practical Implications

As digitization is continuing in the health care sector and point-of-care information may play an increasingly important role in the daily practices of health care professionals, a valid evaluation tool for this kind of medical information is necessary. The usability of such a tool may depend on the user. Tools meant for health care professionals need to be short, whereas tools meant for external organizations that aim to validate information sources may be more comprehensive.

The current situation is problematic: There is no standard, valid tool available for health care professionals for a proper assessment of medical, point-of-care information. The use of the AGREE II instrument for the assessment of clinical guidelines was previously associated with enhanced guideline adoption: increased guideline endorsements, an increase in overall intentions to use guidelines, and an increase in overall quality of guidelines [8]. Therefore, the use of a tool for assessment of point-of-care information may improve the quality and use of this kind of information. Based on the results of this review, we suggest that such a tool should evaluate author-related information and evidence-based methodology. The items from the categories "website quality," "website design and usability," and "website interactivity" can be used to assess whether an information source is truly point-of-care information.

Limitations

When we performed the literature search for this review, we noticed the absence of a common terminology for assessment of trustworthiness of point-of-care information. This is a limitation of this review that might have affected the output of the literature search. A broad search was needed to cover all tools used for different applications in medicine. Similarly, Kwag et al [4] noticed that point-of-care information summaries use different terms. We agree with their statement that a standard definition would be beneficial for the PubMed Mesh vocabulary.

A standard risk-of-bias assessment on each tool could not be performed, as no standard to assess this is currently available. Therefore, it was not possible to distinguish methodologically sound tools from those that are methodologically weak. However, each tool was checked for potential risk of bias in the developmental phase, such as lack of validity and reliability assessment.

Conclusion

In conclusion, this systematic literature review identified 17 different tools for the assessment of the trustworthiness of point-of-care information. These tools encompass a variety of items, but to date, a standard, validated tool is nonexistent. The results of this review may contribute to the development of a standard tool, which may enhance the quality and trustworthiness of point-of-care information in the longer term.

Acknowledgments

The authors would like to acknowledge the Belgian National Institution for Health and Disability Insurance (RIZIV) for funding this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Included tools and criteria.
[DOCX File , 39 KB - jmir_v22i1e15415_app1.docx]

Multimedia Appendix 2
Data summary.
[DOCX File , 37 KB - jmir_v22i1e15415_app2.docx]

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
2. Sadeghi-Bazargani H, Tabrizi JS, Azami-Aghdash S. Barriers to evidence-based medicine: a systematic review. J Eval Clin Pract 2014 Dec;20(6):793-802. [doi: 10.1111/jep.12222] [Medline: 25130323]
3. Banzi R, Liberati A, Moschetti I, Tagliabue L, Moja L. A review of online evidence-based practice point-of-care information summary providers. J Med Internet Res 2010 Jul 07;12(3):e26 [FREE Full text] [doi: 10.2196/jmir.1288] [Medline: 20610379]
4. Kwag KH, González-Lorenzo M, Banzi R, Bonovas S, Moja L. Providing Doctors With High-Quality Information: An Updated Evaluation of Web-Based Point-of-Care Information Summaries. J Med Internet Res 2016 Jan 19;18(1):e15 [FREE Full text] [doi: 10.2196/jmir.5234] [Medline: 26786976]

5. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P, Board of Trustees of the Guidelines International Network. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012 Apr 03;156(7):525-531. [doi: 10.7326/0003-4819-156-7-201204030-00009] [Medline: 22473437]
6. AGREE Next Steps Consortium. 2017. The AGREE II Instrument URL: https://www.agreetrust.org/ [accessed 2019-12-02]
7. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting, and evaluation in health care. Prev Med 2010 Nov;51(5):421-424. [doi: 10.1016/j.ypmed.2010.08.005] [Medline: 20728466]
8. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, AGREE Next Steps Consortium. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. CMAJ 2010 Jul 13;182(10):1045-1052 [FREE Full text] [doi: 10.1503/cmaj.091714] [Medline: 20513780]
9. Higgins J, Lasserson T, Chandler J, Tovey D, Churchill R. Cochrane Community.: The Cochrane Collaboration; 2019 Oct. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates URL: https://community.cochrane.org/mecir-manual [accessed 2019-12-02]
10. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]
11. 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 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
12. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
13. Moher D, Schulz KF, Altman D, CONSORT Group. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. Explore (NY) 2005 Jan;1(1):40-45. [doi: 10.1016/j.explore.2004.11.001] [Medline: 16791967]
14. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Training.: The Cochrane Collaboration; 2019. Cochrane Handbook for Systematic Reviews of Interventions URL: https://training.cochrane.org/handbook/current [accessed 2019-12-02]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
16. Aslani A, Pournik O, Abu-Hanna A, Eslami S. Web-site evaluation tools: a case study in reproductive health information. Stud Health Technol Inform 2014;205:895-899. [Medline: 25160317]
17. Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. Comput Biol Med 1998 Sep;28(5):603-610. [doi: 10.1016/s0010-4825(98)00037-7] [Medline: 9861515]
18. e-Health Ethics Initiative. e-Health Code of Ethics (May 24). J Med Internet Res 2000;2(2):E9 [FREE Full text] [doi: 10.2196/jmir.2.2.e9] [Medline: 11720928]
19. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
20. Kapoun J. Teaching undergrads WEB evaluation: A guide for library instruction. College and Research Libraries News 1998 Jul;59(7):522-523 [FREE Full text]
21. Gillois P, Colombet I, Dréau H, Degoulet P, Chatellier G. A critical appraisal of the use of Internet for calculating cardiovascular risk. Proc AMIA Symp 1999:775-779 [FREE Full text] [Medline: 10566465]
22. Jiang YL. Quality evaluation of orthodontic information on the World Wide Web. Am J Orthod Dentofacial Orthop 2000 Jul;118(1):4-9. [doi: 10.1067/mod.2000.104492] [Medline: 10893465]
23. Pencheon D, Gray M, Melzer D. Oxford Handbook Of Public Health Practice (2nd edition). Oxford: Oxford University Press; 2006.
24. Sandvik H. Health information and interaction on the internet: a survey of female urinary incontinence. BMJ 1999 Jul 03;319(7201):29-32 [FREE Full text] [doi: 10.1136/bmj.319.7201.29] [Medline: 10390457]
25. Robillard JM, Jun JH, Lai J, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
26. Prorok JC, Iserman EC, Wilczynski NL, Haynes RB. The quality, breadth, and timeliness of content updating vary substantially for 10 online medical texts: an analytic survey. J Clin Epidemiol 2012 Dec;65(12):1289-1295. [doi: 10.1016/j.jclinepi.2012.05.003] [Medline: 22974495]

27. Winker MA, Flanagin A, Chi-Lum B, White J, Andrews K, Kennett RL, et al. Guidelines for medical and health information sites on the internet: principles governing AMA web sites. American Medical Association. JAMA 2000;283(12):1600-1606. [doi: 10.1001/jama.283.12.1600] [Medline: 10735398]
28. Delvenne C. Grille d'analyse des sites médicaux sur l'Internet. Liège, Belgium: University of Liège; 1999. URL: http://www.ebm.uliege.be/grille.htm [accessed 2019-12-02]
29. Delvenne C, Pasleau F. Organising access to Evidence-Based Medicine resources on the Web. Comput Methods Programs Biomed 2003 May;71(1):1-10. [doi: 10.1016/s0169-2607(02)00055-x] [Medline: 12725960]
30. Trumble JM, Anderson MJ, Caldwell M, Chuang F, Fulton S, Howard A. Texas Health Science Libraries Consortium. 2006 Nov. A systematic evaluation of evidence based medicine tools for point-of-care URL: http://www.thslc.org/papers.html [accessed 2019-12-02]
31. Yap KY, Raaj S, Chan A. OncoRx-IQ: a tool for quality assessment of online anticancer drug interactions. Int J Qual Health Care 2010 Apr;22(2):93-106. [doi: 10.1093/intqhc/mzq004] [Medline: 20172863]
32. Boyer C, Selby M, Appel RD. The Health On the Net Code of Conduct for medical and health web sites. Stud Health Technol Inform 1998;52 Pt 2:1163-1166. [Medline: 10384641]
33. Joubert M, Aymard S, Fieschi D, Fieschi M. Quality criteria and access characteristics of Web sites: proposal for the design of a health Internet directory. Proc AMIA Symp 1999:824-828 [FREE Full text] [Medline: 10566475]
34. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res 2001;3(4):E28 [FREE Full text] [doi: 10.2196/jmir.3.4.e28] [Medline: 11772543]
35. Moja L, Banzi R. Navigators for medicine: evolution of online point-of-care evidence-based services. Int J Clin Pract 2011 Jan;65(1):6-11. [doi: 10.1111/j.1742-1241.2010.02441.x] [Medline: 21155939]
36. Straus SE, Sackett DL. Using research findings in clinical practice. BMJ 1998 Aug 01;317(7154):339-342 [FREE Full text] [doi: 10.1136/bmj.317.7154.339] [Medline: 9685286]
37. Campbell JM, Umapathysivam K, Xue Y, Lockwood C. Evidence-Based Practice Point-of-Care Resources: A Quantitative Evaluation of Quality, Rigor, and Content. Worldviews Evid Based Nurs 2015 Dec;12(6):313-327. [doi: 10.1111/wvn.12114] [Medline: 26629973]

Abbreviations

AGREE II: Appraisal of Guidelines for Research and Evaluation
AMA: American Medical Association
CART: Completeness, Accuracy, Relevance, Timeliness
eHealth: electronic health
HON: health on the net
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUEST: Quality Evaluation Scoring Tool

Edited by N Zary; submitted 09.07.19; peer-reviewed by M Lorenzo, M Alshehri; comments to author 18.09.19; revised version received 01.10.19; accepted 25.10.19; published 17.01.20.

Please cite as:
Lenaerts G, Bekkering GE, Goossens M, De Coninck L, Delvaux N, Cordyn S, Adriaenssens J, Vankrunkelsven P
Tools to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals: Systematic Review
J Med Internet Res 2020;22(1):e15415
URL: https://www.jmir.org/2020/1/e15415
doi:10.2196/15415
PMID:31951213

Original Paper

Online Information on Electronic Cigarettes: Comparative Study of Relevant Websites From Baidu and Google Search Engines

Ting Chen[1*], PhD; Sarah Gentry[2*], BSBM; Dechao Qiu[1], BMed; Yan Deng[1], BMed; Caitlin Notley[2], PhD; Guangwen Cheng[1], MD; Fujian Song[2], PhD

[1]School of Public Health, Hubei Provincial Key Laboratory of Occupational Hazard Identification & Control, Wuhan University of Science & Technology, Wuhan, China

[2]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Fujian Song, PhD
Norwich Medical School
University of East Anglia
Norwich, NR4 7TJ
United Kingdom
Phone: 44 1603 591253
Email: fujian.song@uea.ac.uk

Abstract

Background: Online information on electronic cigarettes (e-cigarettes) may influence people's perception and use of e-cigarettes. Websites with information on e-cigarettes in the Chinese language have not been systematically assessed.

Objective: The aim of this study was to assess and compare the types and credibility of Web-based information on e-cigarettes identified from Google (in English) and Baidu (in Chinese) search engines.

Methods: We used the keywords *vaping* or *e-cigarettes* to conduct a search on Google and the equivalent Chinese characters for Baidu. The first 50 unique and relevant websites from each of the two search engines were included in this analysis. The main characteristics of the websites, credibility of the websites, and claims made on the included websites were systematically assessed and compared.

Results: Compared with websites on Google, more websites on Baidu were owned by manufacturers or retailers (15/50, 30% vs 33/50, 66%; $P<.001$). None of the Baidu websites, compared to 24% (12/50) of Google websites, were provided by public or health professional institutions. The Baidu websites were more likely to contain e-cigarette advertising ($P<.001$) and less likely to provide information on health education ($P<.001$). The overall credibility of the included Baidu websites was lower than that of the Google websites ($P<.001$). An age restriction warning was shown on all advertising websites from Google (15/15) but only on 10 of the 33 (30%) advertising websites from Baidu ($P<.001$). Conflicting or unclear health and social claims were common on the included websites.

Conclusions: Although conflicting or unclear claims on e-cigarettes were common on websites from both Baidu and Google search engines, there was a lack of online information from public health authorities in China. Unbiased information and evidence-based recommendations on e-cigarettes should be provided by public health authorities to help the public make informed decisions regarding the use of e-cigarettes.

(J Med Internet Res 2020;22(1):e14725) doi:10.2196/14725

KEYWORDS

electronic nicotine delivery system; electronic cigarette; online health information; internet-based information

Introduction

Electronic cigarettes (e-cigarettes) are also called electronic nicotine delivery systems or nicotine vaping products, although nicotine-free e-cigarettes exist. Users of e-cigarettes inhale a vapor that may contain nicotine by heating a solution with battery power. Since the first patent of e-cigarettes in 2003 in China, different types of e-cigarettes have been developed [1], and 80% of the global e-cigarette products are manufactured in China [2]. More recently, *heat-not-burn* tobacco (HNBT)

products have been developed to electronically heat processed tobacco to produce aerosol without combustion [3], and they have been increasingly used in Japan and many other countries [4].

Along with dramatically increased popularity, the potential impact of e-cigarettes on public health remains controversial [5]. Available research evidence indicates that the use of e-cigarettes is considerably less harmful than the use of traditional combustible cigarettes; hence, e-cigarettes have been recommended for smokers as cessation aids or harm reduction alternatives [6]. However, some researchers or tobacco control experts remain skeptical about the impact of e-cigarettes on public health for the following reasons. First, e-cigarettes are not risk free, and there is limited evidence on the effects of e-cigarettes for smoking cessation [7]. Second, possible *gateway effects* may increase smoking initiation among young people [8,9]. E-cigarettes may be categorized differently as tobacco products, therapeutic products, or consumer products by authorities in different countries [10]. Although the sale and use of e-cigarettes are unregulated in many countries (eg, mainland China), they have been banned for sale in some countries (eg, Australia and Brazil) but encouraged for smoking cessation or harm reduction in others (eg, United Kingdom) [11,12].

There are very different prevalence rates of use of e-cigarettes across countries [5]. The prevalence of current use of e-cigarettes in 2018 was 6.3% in England [13] and only 0.9% in China [14]. The high prevalence of current use of e-cigarettes in the United Kingdom could be explained by various public health and health organizations' recommendation of e-cigarettes as cessation aids or less harmful alternatives for smokers. Nonetheless, the very low prevalence of current use of e-cigarettes in China is noteworthy, which is contrasting to a high prevalence of cigarette smoking, largely unregulated e-cigarette marketing, and substantial amount of global e-cigarette products made in China [2].

Online advertising is an important approach for manufacturers and retailers to marketing e-cigarettes and accessories. A systematic review found an association between exposure to online marketing and intention to use or trial use of e-cigarettes [15]. Furthermore, the internet remains an important source of health information for the general public, and the increased availability of mobile phones has facilitated the internet access globally [16]. A study found that 80% of those studied used internet search engines to look for information on e-cigarettes, and online information may influence people's perception and use of e-cigarettes [17].

A study evaluated websites of e-cigarette manufacturers in China and concluded that e-cigarette marketing should be better regulated [18]. However, online information on e-cigarettes in Chinese has not been systematically assessed. This study aimed to assess and compare types and credibility of Web-based information on e-cigarettes identified from Google (in English language) and Baidu (in Chinese language) search engines.

Methods

Search for and Inclusion of Relevant Websites

We found no studies on keywords used by ordinary users to search for online information on e-cigarettes. The term *electronic cigarette* was used in a study of websites of e-cigarette manufacturers in China [18]. Therefore, we used keywords *vaping* or *e-cigarettes* for Google search engine and equivalent Chinese characters for Baidu search engine. The Google search was conducted in England, and the Baidu search was conducted in Wuhan, China, on January 21, 2019. The Google search provided approximately 19 million results, and the Baidu search showed approximately 37 million results. Considering that typical internet users would focus on a limited number of Web search pages and restrictions of available time and other resources, we selected the first 50 unique and relevant websites from each of the 2 search engines, after excluding duplicates, irrelevant websites, and those that were no longer active.

Data Extraction and Analysis

We developed a data extraction form, which was tested and revised using the first 4 included websites from each of the 2 search engines (Multimedia Appendix 1). Then, 2 authors independently extracted information from included websites (TC, YD, and GC from Baidu websites and SG and DQ from Google websites). To ensure consistency in data extraction, a third author (FS) compared and checked all extracted data and made an arbitrary decision for any disagreements between reviewers.

We extracted the following data from the relevant websites: characteristics of the website, main messages regarding the use of e-cigarettes, and website credibility (Multimedia Appendix 1). Claims regarding e-cigarettes were categorized as smoking cessation claims, health claims, social claims, and age restriction warnings, according to the method used by Hsu et al [1]. From e-cigarette advertising websites, we collected data on types of e-cigarettes, including Cigalike (closed system), eGo (open system), and Mods (advanced personal vaporizers). We also recorded whether HNBT products were mentioned, although HNBT are products different from e-cigarettes.

Quality of included websites was assessed using the Quality Evaluation Scoring Tool (QUEST) [19]. The QUEST has been validated in a previous study and can be used to assess the quality of online information from the following 6 aspects: authorship, attribution, conflict of interest, currency, complementarity, and tone of claims. We removed the *complementarity* item for patient-physician relationship from the QUEST tool because it is of limited relevance in this study. In addition, attribution (evidence source) was revised to distinguish single and multiple studies and systematic reviews. For a website, the quality score ranges from 0 to 28, in which a higher score indicates a better website quality (see Multimedia Appendix 1 for more details).

We summarized data extracted from included websites in tables and reported percentages of the main website characteristics. Differences in the proportion of websites with certain

characteristics between Google and Baidu search engines were tested using Pearson chi-square test or Fisher exact test if any of the expected cell sizes were less than 5. Two-sample Wilcoxon rank-sum method was used to test the difference in the total QUEST score between websites from Google and Baidu search engines. Statistical significance was defined as 2-sided $P \leq .05$. Stata/Special Edition 14.2 (StataCorp LLC, USA) software was used for statistical analysis.

Results

Characteristics of Included Websites

The main characteristics of included websites are shown in Table 1. All the included websites from Baidu search engine were located in mainland China. For the websites from Google search engine, 24 were located in the United States, 20 in the United Kingdom, and 6 in other countries (including 1 each in Australia and Canada, 2 in New Zealand, and 2 with unclear locations). Providers of the included websites from Baidu search engine were mostly e-cigarette manufacturers/retailers (33/50, 66%) and mass media or information technology (IT) services (15/50, 30%). None of the included websites from Baidu search engine were owned by public or health professional institutions. For the 50 websites from Google search engine, 12 (24%) were owned by public or health professional institutions, 15 (30%) by e-cigarette manufacturers or retailers, 12 (24%) by media or IT companies, and 11 (22%) by charity or not-for-profit nongovernment organizations (Table 1).

Compared with websites from Google search engine, those from Baidu search engine were more likely to contain e-cigarette advertising (37/50, 74% vs 15/50, 30%; $P<.001$) and user feedback (21/50, 42% vs 11/50, 22%; $P=.03$), but they were much less likely to provide information on health education (3/50, 6% vs 30/50, 60%; $P<.001$). In addition, relatively more Baidu websites provided e-cigarette–related news, although the difference in the proportion was statistically nonsignificant ($P=.14$). Contents covered by the included websites were similar from the 2 search engines, including descriptions about what are e-cigarettes, e-cigarettes' harm or risk, and the role of e-cigarettes for smoking cessation (Table 1). However, websites from Baidu search engine were less likely to consider regulation issues than those from Google search engine ($P=.04$).

Table 1. The main characteristics of included websites.

Characteristics	Baidu (N=50), n (%)	Google (N=50), n (%)	*P* value
Country			__[a]
China	50 (100)	0 (0)	
United States	0 (0)	24 (48)	
United Kingdom	0 (0)	20 (40)	
Other	0 (0)	6 (12)	
Website owner[b]			<.001
Public or health professional institutions	0 (0)	12 (24)	
Manufacturer/retailer	33 (66)	15 (30)	
Media/information technology services	15 (30)	12 (24)	
Charity/not-for-profit nongovernment organization/other	2 (4)	11 (22)	
Type of information[c]			
Health education	3 (6)	30 (60)	<.001
News	21 (42)	14 (28)	.14
Advertisement	37 (74)	15 (30)	<.001
Blogs/user feedback	21 (42)	11 (22)	.03
Content coverage[c]			
What are e-cigarettes[d]	31 (62)	35 (70)	.40
E-cigarettes' harm	32 (64)	40 (80)	.08
E-cigarettes for quitting	35 (70)	31 (62)	.40
E-cigarette regulation	23 (46)	33 (66)	.04

[a]Not applicable.

[b]*P* value for the website owner was a test of null hypothesis that there was no statistically significant difference in the proportion of website owners between Baidu and Google search engines.

[c]As individual items were not mutually exclusive for type of information and content coverage, *P* values for items belonging to these 2 variables were tests of null hypothesis that there was no statistically significant difference in the proportion of an individual item between the 2 search engines.

[d]e-cigarettes: electronic cigarettes.

Quality of Included Websites

The included websites from Baidu search engine had lower modified QUEST scores than those from Google search engine. The median QUEST score was 5 (range: 0-16) for the Baidu websites and 15 (range: 0-27) for the Google websites (P<.001). After excluding websites owned by manufacturers or retailers, the median QUEST score was 6.3 (range: 2-11) for the Baidu websites and 18.8 (range: 9-27) for the Google websites (P<.001). Baidu websites tended to be more current (ie, more recently updated) than Google websites (Table 2). However, compared with Google websites, Baidu websites were associated with a higher proportion of no indication of authorship, lacking or inadequate scientific evidence, high risk of conflict of interest, and fully supportive tone of claims (Table 2). For a subgroup analysis of websites not owned by manufacturers or retailers, the differences in quality between Baidu and Google websites were increased for attribution, high risk of conflict of interest, and tone of claims (Table 2). It is noteworthy that information was considered to be unbiased for 20 of the 35 Google websites that were not owned by manufacturers or retailers, whereas information was of high or unclear risk of bias for all of the Baidu websites (Table 2).

Table 2. Modified Quality Evaluation Scoring Tool website quality criteria and scores. For categorical variables with items that were mutually exclusive, a single *P* value was obtained from chi-square test. *P* values for the Quality Evaluation Scoring Tool score were based on Wilcoxon rank-sum test.

Criteria	All included websites			Nonmanufacturers or nonretailers		
	Baidu (N=50), n (%)	Google (N=50), n (%)	*P* value	Baidu (N=17), n (%)	Google (N=35), n (%)	*P* value
Authorship			.008			.13
No indication of authorship	44 (88)	33 (66)		12 (71)	18 (51)	
All other indications of authorship	6 (12)	10 (20)		5 (29)	10 (29)	
Author/qualification clearly stated	0 (0)	7 (14)		0 (0%)	7 (20)	
Attribution—a			<.001			<.001
No sources	28 (56)	9 (18)		10 (59)	0 (0)	
Mention of expert source and research findings, but insufficiently	20 (40)	14 (28)		7 (41)	12 (34)	
Reference to at least one identifiable scientific study	2 (4)	10 (20)		0 (0)	6 (17)	
Reference to mainly identifiable scientific studies	0 (0)	17 (34)		0 (0)	17 (49)	
Attribution—b			<.001			.001
Not available; in vitro, animal, and editorials	46 (92)	23 (46)		15 (88)	14 (40)	
Single journal article	3 (6)	2 (4)		2 (12)	1 (3)	
Multiple journal articles	1 (2)	15 (30)		0 (0)	13 (37)	
Systematic reviews of studies	0 (0)	10 (20)		0 (0)	7 (20)	
Conflicts of interest			<.001			<.001
High risk of conflict of interest	27 (54)	15 (30)		5 (29)	1 (3)	
Unclear risk of conflict of interest	23 (46)	15 (30)		12 (71)	14 (40)	
Unbiased information	0 (0)	20 (40)		0 (0)	20 (57)	
Currency			.001			.006
No date present	7 (14)	25 (50)		0 (0)	12 (34)	
Dated but 1 year old or older	7 (14)	3 (6)		0 (0)	3 (9)	
Dated within the last 1 year	36 (72)	22 (44)		17 (100)	20 (57)	
Tone of claims			<.001			<.001
Fully supported	36 (72)	12 (24)		14 (82)	3 (9)	
Mainly supported	13 (26)	17 (34)		3 (18)	12 (34)	
Balanced/cautious support	1 (2)	21 (42)		0 (0)	20 (57)	

Claims on the Included Websites

Table 3 shows claims or messages on the included websites, according to whether websites were owned by manufacturers or retailers. For websites owned by manufacturers/retailers, Baidu websites were more likely to claim that the use of e-cigarettes helped with smoking cessation (22/33, 67% vs 5/15, 33%) and that e-cigarettes were less harmful than combustible cigarettes (20/33, 61% vs 7/15, 47%), although the differences between the 2 search engines were statistically nonsignificant

(*P*>.05). Google websites owned by manufacturers/retailers were more likely to claim that the use of e-cigarettes was cheaper than the use of combustible cigarettes (7/15, 47% vs 3/33, 9%; *P*=.006). An age restriction warning was shown on all Google websites owned by manufacturers or retailers but by only on 10 of the 33 (30%) Baidu websites owned by manufacturers or retailers (*P*<.001).

After excluding websites owned by manufacturers or retailers, the differences in quitting, health claims, and social claims between Baidu and Google websites were statistically nonsignificant (Table 3). However, 3 of the 17 Baidu websites that were not owned by manufacturers or retailers claimed that e-cigarettes were more harmful than combustible cigarettes, compared with only 1 of the 35 such Google websites (*P*=.21).

Table 3. Claims or messages from the included websites.

Claims	Manufacturer/retailer websites			Nonmanufacturer or nonretailers		
	Baidu (N=33), n (%)	Google (N=15), n (%)	*P* value	Baidu (N=17), n (%)	Google (N=35), n (%)	*P* value
Electronic cigarette for quitting[a]						
Help quit	22 (67)	5 (33)	.053	8 (47)	14 (40)	.86
Not help quit	1 (3)	0 (0)	—[b]	2 (12)	7 (20)	—
Unclear/other	10 (30)	10 (67)	—	7 (41)	14 (40)	—
Health claims[a]						
Healthier than cigarettes	20 (61)	7 (47)	.53	7 (41)	16 (46)	.21
More harmful than cigarettes	0 (0)	0 (0)	—	3 (18)	1 (3)	—
Unclear/other	13 (39)	8 (53)	—	7 (41)	18 (51)	—
Social claims (multiple choices allowed)[c]						
Less expensive than cigarettes	3 (9)	7 (47)	.006	0 (0)	2 (6)	>.99
Cleaner than cigarettes	14 (42)	4 (27)	.35	4 (24)	4 (11)	.41
More socially acceptable	5 (15)	3 (20)	.69	1 (6)	6 (17)	.40
Age claims/warning [a]						
Yes	10 (30)	15 (100)	<.001	2 (12)	29 (83)	<.001
No	23 (70)	0 (0)	—	15 (88)	6 (17)	—

[a]When items belonging to a variable were mutually exclusive, a single *P* value was obtained from chi-square test for the variable (ie, e-cigarette for quitting, health claims, and age claims).

[b]Not applicable.

[c]If items for a variable were not mutually exclusive (ie, social claims), a *P* value was shown for each row from a test of null hypothesis that there was no statistically significant difference in the proportion of the item between the 2 search engines.

Types of Electronic Cigarettes Advertised

Table 4 shows types of e-cigarettes advertised on websites from Baidu and Google search engines. Relatively more Google websites were advertising first-generation e-cigarettes than Baidu websites, although the difference was statistically nonsignificant (*P*=.07). The second- and third-generation types of e-cigarettes were similarly promoted on most advertising websites from both Baidu and Google search engines. However, HNBT products were advertised in much more Baidu websites than Google websites (24/37, 65% vs 1/15, 7%; *P*<.001).

Table 4. Types of electronic cigarettes on advertising websites.

Types of electronic cigarettes	Baidu (N=37), n (%)	Google (N=15), n (%)	*P* value
Cigalike	14 (38)	10 (67)	.07
Advanced personal vaporizers or eGo	27 (73)	14 (93)	.15
Mods	30 (81)	13 (87)	.63
Heat-not-burn tobacco	24 (65)	1 (7)	<.001

Discussion

Prinicpal Findings

Findings of this study revealed that the overall credibility and quality of the included Baidu websites were much lower than those of the included Google websites. Compared with Google websites, the included Baidu websites were more likely to be owned by manufacturers or retailers, more likely to advertise HNBT products, less likely to focus on health education, and less likely to have an age restriction warning. The included websites from the 2 search engines similarly provided conflicting or unclear claims regarding whether the use of e-cigarettes helped smoking cessation and whether the use of e-cigarettes was more or less harmful and more or less socially acceptable compared with the use of conventional cigarettes.

Claims made on the marketing websites included in this study were similar to those identified by previous studies, including claims that e-cigarettes could help smoking cessation and that e-cigarettes are safer, cheaper, cleaner, and more socially acceptable than combustible cigarettes [15,18,20,21]. However, conflicting or unclear health and social claims about e-cigarettes were made on nonmanufacturer/retailer websites from both Baidu and Google search engines. The public may trust online information from public health authorities more than the online information from manufacturers and retailers. None of public health authority websites were identified from Baidu search engine. From Google search engine, 3 public health authorities' websites (1 each from Canada [22], the United States [23], and the United Kingdom [24]) clearly support the use of e-cigarettes for smoking cessation and indicate that the use of e-cigarettes was less harmful than cigarette smoking. According to information on 1 of the included Baidu websites (Baidu Wikipedia website on e-cigarettes, endorsed by Chinese Society of Preventive Medicine [25]), the effect of e-cigarettes for smoking cessation was uncertain and the use of e-cigarettes was more harmful than cigarette smoking. Although many advertising websites from Baidu search engine had positive claims on the use of e-cigarettes, there were no credible websites in Chinese language that supported the use of e-cigarettes for smoking cessation or as a safer alternative to cigarette smoking. Further research is required to investigate whether this is a reason for a very low uptake of e-cigarettes at 0.9% in 2018 in China [14], compared with 6.3% in the United Kingdom [13].

To avoid *gateway effects* of e-cigarette use among young people, some tobacco control experts and public health agencies may stress potential harms of e-cigarettes, and they have been reluctant to support the use of e-cigarettes for smoking cessation or harm reduction [9]. Due to a lack of unbiased information from credible sources, young people in China may be more exposed to marketing information from manufacturers and retailers. E-cigarettes are often displayed as fashion or health products with modern stylish designs to attract young users [26]. Although government agencies in China ban selling of e-cigarettes to underaged young people [27], a large proportion of marketing websites from Baidu search engine had no age restriction warnings.

A public health challenge in China is the very high prevalence of cigarette smoking among men (52.1% in 2015 and 50.5% in 2018) [14]. In addition to other tobacco control measures, it is important to investigate whether the use of e-cigarettes could increase smoking cessation in China, for example, as has being observed in the United Kingdom [6]. However, further research is required to comprehensively understand the public health impact of e-cigarettes on smoking initiation in young people, cessation among current smokers, and changes in all related diseases. Currently, there is an urgent need for unbiased information and evidence-based recommendations on e-cigarettes from public health authorities in China.

Strengths and Limitations

This is the first study to compare information on e-cigarettes between Baidu (in Chinese language) and Google (in English language) search engines. The credibility of and claims on the included websites were systematically assessed. Although the existing study had focused on advertising or marketing websites [18], this study included various websites with information on e-cigarettes.

We included only the first 50 websites from each of the 2 search engines, assuming that ordinary internet users tend to look at the first few search pages. This study considered only websites in Chinese and English language. We used simple terms (*e-cigarette* or *vaping*) to identify relevant websites, although internet users might use or add other search terms (eg, harm, risk, and quitting). Therefore, this study may have missed some influential websites.

Conclusions

The overall quality of the included Baidu websites was much lower than that of the included Google websites, although conflicting or unclear claims on e-cigarettes were common on websites from both Baidu and Google search engines. Compared with websites from Google search engine, relatively more websites from Baidu search engine were e-cigarette advertising. Particularly, there was no credible information on e-cigarettes from public health authorities in China. Unbiased information and evidence-based recommendations on e-cigarettes should be provided by public health authorities to help the public make informed decisions regarding the use of e-cigarettes.

Authors' Contributions

FS designed the study. FS, SG, TC, and CN contributed to the development of data extraction form. FS conducted Google search, and TC conducted Baidu search. FS, TC, SG, DQ, CN, GC, and FS collected data from the included websites. FS analyzed data and drafted the manuscript. All authors commented on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data extraction form: assessment of Web-based information on electronic cigarettes.
[PDF File (Adobe PDF File), 453 KB - jmir_v22i1e14725_app1.pdf]

References

1. Hsu G, Sun JY, Zhu SH. Evolution of electronic cigarette brands from 2013-2014 to 2016-2017: analysis of brand websites. J Med Internet Res 2018 Mar 12;20(3):e80 [FREE Full text] [doi: 10.2196/jmir.8550] [Medline: 29530840]
2. Xu X, Wang X, Zhang X, Liu Y, He H, Mackay J. The debate on regulation of e-cigarettes in China. Lancet Respir Med 2016 Nov;4(11):856-858. [doi: 10.1016/S2213-2600(16)30313-7] [Medline: 27743868]
3. Farsalinos KE, Yannovits N, Sarri T, Voudris V, Poulas K. Nicotine delivery to the aerosol of a heat-not-burn tobacco product: comparison with a tobacco cigarette and e-cigarettes. Nicotine Tob Res 2018 Jul 09;20(8):1004-1009. [doi: 10.1093/ntr/ntx138] [Medline: 28637344]
4. Tabuchi T, Gallus S, Shinozaki T, Nakaya T, Kunugita N, Colwell B. Heat-not-burn tobacco product use in Japan: its prevalence, predictors and perceived symptoms from exposure to secondhand heat-not-burn tobacco aerosol. Tob Control 2018 Jul;27(e1):e25-e33 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053947] [Medline: 29248896]
5. Gravely S, Driezen P, Ouimet J, Quah AC, Cummings KM, Thompson ME, et al. Prevalence of awareness, ever-use and current use of nicotine vaping products (NVPs) among adult current smokers and ex-smokers in 14 countries with differing regulations on sales and marketing of NVPs: cross-sectional findings from the ITC Project. Addiction 2019 Jun;114(6):1060-1073 [FREE Full text] [doi: 10.1111/add.14558] [Medline: 30681215]
6. McNeill A, Brose L, Calder R, Bauld L, Robson D. Public Health England. London; 2018. Evidence review of e-cigarettes and heated tobacco products URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684963/Evidence_review_of_e-cigarettes_and_heated_tobacco_products_2018.pdf
7. Hajek P, Phillips-Waller A, Przulj D, Pesola F, Myers Smith K, Bisal N, et al. A randomized trial of e-cigarettes versus nicotine-replacement therapy. N Engl J Med 2019 Feb 14;380(7):629-637. [doi: 10.1056/NEJMoa1808779] [Medline: 30699054]
8. Glantz SA, Bareham DW. E-cigarettes: use, effects on smoking, risks, and policy implications. Annu Rev Public Health 2018 Apr 01;39:215-235 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-013757] [Medline: 29323609]
9. Stanbrook MB. Electronic cigarettes and youth: a gateway that must be shut. Can Med Assoc J 2016 Aug 09;188(11):785 [FREE Full text] [doi: 10.1503/cmaj.160728] [Medline: 27431301]
10. Electronic Nicotine Delivery Systems. In: Pediatrics. Delhi: The World Health Organization; Oct 26, 2015:1018-1026.
11. Shapiro H. No Fire, No Smoke: The Global State of Tobacco Harm Reduction. London: Knowledge Action Change; 2018.
12. Kennedy RD, Awopegba A, De León E, Cohen JE. Global approaches to regulating electronic cigarettes. Tob Control 2017 Jul;26(4):440-445 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053179] [Medline: 27903958]
13. ONS. Statistics on Smoking England. In: NHS Digital, the Government Statistical Service. London: The Government Statistical Service; 2019:2019.
14. Report of 2018 Adult Smoking Survey in China. Beijing: Chinese Center for Disease Prevention and Control; 2018. URL: https://www.tobaccofreekids.org/assets/global/pdfs/en/GATS_China_2018_FactSheet.pdf
15. Collins L, Glasser AM, Abudayyeh H, Pearson JL, Villanti AC. E-cigarette marketing and communication: how e-cigarette companies market e-cigarettes and the public engages with e-cigarette information. Nicotine Tob Res 2019 Jan 01;21(1):14-24 [FREE Full text] [doi: 10.1093/ntr/ntx284] [Medline: 29315420]
16. Tonsaker T, Bartlett G, Trpkov C. Health information on the internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
17. Emery SL, Vera L, Huang J, Szczypka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. Tob Control 2014 Jul;23(Suppl 3):iii17-iii25 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051648] [Medline: 24935893]
18. Yao T, Jiang N, Grana R, Ling PM, Glantz SA. A content analysis of electronic cigarette manufacturer websites in China. Tob Control 2014 Oct 21;25(2):188-194. [doi: 10.1136/tobaccocontrol-2014-051840] [Medline: 4405407]
19. Robillard JM, Jun JH, Lai JA, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
20. Klein EG, Berman M, Hemmerich N, Carlson C, Htut S, Slater M. Online E-cigarette marketing claims: a systematic content and legal analysis. Tob Regul Sci 2016 Jul;2(3):252-262 [FREE Full text] [doi: 10.18001/TRS.2.3.5] [Medline: 27446984]
21. McCausland K, Maycock B, Leaver T, Jancey J. The messages presented in electronic cigarette-related social media promotions and discussion: scoping review. J Med Internet Res 2019 Feb 05;21(2):e11953 [FREE Full text] [doi: 10.2196/11953] [Medline: 30720440]

22. Government of Canada. About vaping URL: https://www.canada.ca/en/health-canada/services/smoking-tobacco/vaping.html
23. Electronic Cigarettes. 2019. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/index.htm
24. National Health Service, UK. E-cigarettes / vapes URL: https://www.nhs.uk/smokefree/help-and-advice/e-cigarettes
25. Baidu Baike. 2019. Electronic cigarettes URL: https://baike.baidu.com/item/???/683778?fr=aladdin
26. Jiang N, Ho SY, Lam TH. Electronic cigarette marketing tactics in mainland China. Tob Control 2017 Mar;26(2):230-232. [doi: 10.1136/tobaccocontrol-2015-052824] [Medline: 27071732]
27. SAMR-STMA. State Administration for Market Regulation and State Tobacco Monoply Administration. 2018 Aug 31. Banning sales of electronic cigarettes to under aged young people -Notification of the State Administration for Market Regulation and State Tobacco Monoply Administration (31-08-2018) URL: http://www.tobacco.gov.cn/html/49/86870396_n.html [accessed 2019-04-24]

Abbreviations

e-cigarette: electronic cigarette
HNBT: heat-not-burn tobacco
IT: information technology
QUEST: Quality Evaluation Scoring Tool

Edited by G Eysenbach; submitted 16.05.19; peer-reviewed by K McCausland, M Tkác; comments to author 28.09.19; revised version received 16.10.19; accepted 19.12.19; published 24.01.20.

Please cite as:
Chen T, Gentry S, Qiu D, Deng Y, Notley C, Cheng G, Song F
Online Information on Electronic Cigarettes: Comparative Study of Relevant Websites From Baidu and Google Search Engines
J Med Internet Res 2020;22(1):e14725
URL: http://www.jmir.org/2020/1/e14725/
doi:10.2196/14725
PMID:32012069

Original Paper

Analysis of Collective Human Intelligence for Diagnosis of Pigmented Skin Lesions Harnessed by Gamification Via a Web-Based Training Platform: Simulation Reader Study

Christoph Rinner[1], PhD; Harald Kittler[2], MD; Cliff Rosendahl[3], MBBS, PhD; Philipp Tschandl[2], MD, PhD

[1]Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[2]Department of Dermatology, Medical University of Vienna, Vienna, Austria

[3]Faculty of Medicine, University of Queensland, Brisbane, Australia

Corresponding Author:
Philipp Tschandl, MD, PhD
Department of Dermatology
Medical University of Vienna
Währinger Gürtel 18-20
Vienna, 1090
Austria
Phone: 43 40400 ext 77000
Email: philipp.tschandl@meduniwien.ac.at

Abstract

Background: The diagnosis of pigmented skin lesion is error prone and requires domain-specific expertise, which is not readily available in many parts of the world. Collective intelligence could potentially decrease the error rates of nonexperts.

Objective: The aim of this study was to evaluate the feasibility and impact of collective intelligence for the detection of skin cancer.

Methods: We created a gamified study platform on a stack of established Web technologies and presented 4216 dermatoscopic images of the most common benign and malignant pigmented skin lesions to 1245 human raters with different levels of experience. Raters were recruited via scientific meetings, mailing lists, and social media posts. Education was self-declared, and domain-specific experience was tested by screening tests. In the target test, the readers had to assign 30 dermatoscopic images to 1 of the 7 disease categories. The readers could repeat the test with different lesions at their own discretion. Collective human intelligence was achieved by sampling answers from multiple readers. The disease category with most votes was regarded as the collective vote per image.

Results: We collected 111,019 single ratings, with a mean of 25.2 (SD 18.5) ratings per image. As single raters, nonexperts achieved a lower mean accuracy (58.6%) than experts (68.4%; mean difference=−9.4%; 95% CI −10.74% to −8.1%; P<.001). Collectives of nonexperts achieved higher accuracies than single raters, and the improvement increased with the size of the collective. A collective of 4 nonexperts surpassed single nonexperts in accuracy by 6.3% (95% CI 6.1% to 6.6%; P<.001). The accuracy of a collective of 8 nonexperts was 9.7% higher (95% CI 9.5% to 10.29%; P<.001) than that of single nonexperts, an improvement similar to single experts (P=.73). The sensitivity for malignant images increased for nonexperts (66.3% to 77.6%) and experts (64.6% to 79.4%) for answers given faster than the intrarater mean.

Conclusions: A high number of raters can be attracted by elements of gamification and Web-based marketing via mailing lists and social media. Nonexperts increase their accuracy to expert level when acting as a collective, and faster answers correspond to higher accuracy. This information could be useful in a teledermatology setting.

(J Med Internet Res 2020;22(1):e15597) doi:10.2196/15597

KEYWORDS

skin cancer; crowdsourcing; games, experimental; diagnosis; melanoma; nevi; skin pigmentation; basal cell carcinoma; dermatoscopy

Introduction

Background

Accurate diagnosis of pigmented skin lesions requires experience and depends on the availability of specifically trained physicians [1]. The introduction of convolutional neural networks boosted the development of decision support systems that diagnose pigmented skin lesions independent of human expertise [2]. Recently, it has been shown that computer algorithms outperform humans in the diagnosis of most types of pigmented skin lesions, including melanoma (MEL) [3]. Despite their good performance, computer algorithms are not well accepted in the medical community, probably because of the lack of interpretability of the output, suboptimal human-machine interfaces, and insufficient integration in the clinical workflow. Another limitation of algorithms is their decreased performance for out-of-distribution images [4], which impairs their generalizability. Beside machine learning algorithms, other approaches exist that are directed toward delivering expert dermatologic service for the accurate diagnosis of pigmented skin lesions. Store-and-forward telemedicine technologies are well tested for triaging [5,6], but are especially useful in regions where specialist service is not readily available [7]. Most telemedical applications use dermatoscopic images, as dermatoscopy improves the diagnosis of pigmented skin lesions in comparison with examination with the unaided eye [8].

Aside the help of computer algorithms, the use of collective intelligence could improve the accuracy of diagnosis to, or even beyond, the level of experts. Collective intelligence has emerged in the last two decades with the rise of interconnected groups of people and computers, collectively solving difficult tasks [9,10]. In radiology, for example, artificial swarm intelligence has been used to increase the diagnostic accuracy when reviewing chest x-rays for the presence of pneumonia [11].

Objectives

To measure the effect of swarm intelligence for the diagnosis of pigmented skin lesions, we created a publicly available dataset of 10,015 dermatoscopic images and developed a Web-based study platform with elements of gamification. We also aimed at providing a publicly available benchmark dataset with human diagnoses and corresponding metadata that will be helpful for developing and testing future machine learning algorithms.

The aim of this study was to find out whether a collective of nonexperts can reach expert-level (ie, being frequently consulted by experts and advanced users) accuracy in diagnosing skin cancer on dermatoscopic images and to find out the collective size needed for this.

Methods

Web-Based Training Platform

The Web-based platform DermaChallenge [12], which was developed at the Medical University of Vienna, is an interactive training platform to educate dermatologists and other physicians interested in dermatoscopy via gamification and individual feedback. The platform is split into the back end and the front end, and both are deployed on a stack of well-known Web technologies (Linux, Apache, MySQL, and PHP) at the Medical University of Vienna. The back end is programmed using Laravel (version 5.5) [13] and offers a Representational State Transfer interface to load and persist data as well as JavaScript Object Notation Web tokens to authenticate participants. To protect participants' data, the Transport Layer Security and Secure Sockets Layer protocol are used to encrypt all communications. The front end is a React (version 16) [14] app optimized for mobile devices (mobile phones and tablets) but can also be used on any other platform via a JavaScript-enabled Web browser. The user interface is based on Semantic UI React [15] components and Redux [16] to manage the state of the application. Before public deployment, 5 users tested the platform.

To participate in the study, participants had to register with a username, a valid email address, and a password. In addition, we asked participants their age (age groups spanning 10 years), gender, country, profession, and years of experience in dermatoscopy. Ordered options for the last item were (1) less than 1 year, (2) opportunistic use for more than 1 year, (3) regular use for 1 to 5 years, (4) regular use for more than 5 years, or (5) more than 10 years of experience. Groups 1 to 4 were regarded as *nonexperts* and group 5 as *experts*. The training platform was publicly available; to start playing, only registration had to be completed and the email address had to be verified.

For gamification, the training platform is structured into stepwise levels with different tasks of varying degrees of difficulty. As the platform is available to any user with a valid email address, we needed to verify plausibility of self-declared experience status. For this, the first 3 levels were *screening levels* and comprised simple domain-specific tasks (assign 1 of the 7 possible diagnoses to 10 cases, separate MELs from non-MELs, and separate seborrheic keratoses from other lesions) to introduce the platform and to assess the basic skills of the participants (see Tschandl et al [3]). The target level (level 4) was available only after the completion of 1 round in each screening level (levels 1-3). In the target level, 30 dermatoscopic images were presented to the participants. The images in this level were taken from a master dataset of 10,015 dermatoscopic images (the Human Against Machine with 10000 training images (HAM10000) dataset [17], publicly available at The Harvard Dataverse [18] and the ISIC-Archive [19]), where all malignant diagnoses were verified via histopathology, and ground truth for benign cases was distributed in a similar fashion to the source dataset. The task of the raters in this level was to select the correct diagnosis out of 7 predefined categories: (1) actinic keratosis/intraepithelial carcinoma (AKIEC), (2) basal cell carcinoma (BCC), (3) seborrheic keratosis/solar lentigo/lichen planus–like keratosis (*benign keratinocytic lesions*, BKL), (4) dermatofibroma (DF), (5) MEL, (6) nevus (NV), and (7) vascular lesions (VASC). In clinical practice, more than 95% of pigmented skin lesions will fall into 1 of the 7 categories [8]. The batches of 30 images per round had a predefined composition of diseases (3×AKIEC, 4×BCC,

4×BKL, 3×DF, 5×MEL, 9×NV, and 2×VASC). Cases were selected randomly from each disease category according to this blueprint. To ensure a balanced distribution of NV, this category was stratified into 3 groups according to ground truth as published with the HAM10000 dataset (histopathology, follow-up with digital dermatoscopy, and expert consensus [17]). Each batch of NV included 3 cases of each category. In the analysis of this paper, only data from these 4 levels are used.

Raters were allowed to play more than 1 round per level. When a round was completed, raters were able to see their scoring rank in an *all-time* or *current month* leaderboard for each level and could compare their accuracy with others. To avoid cheating, each image was shown for a maximum of 25 seconds. After 20 seconds, raters received a warning that the system will continue to the next case in 5 seconds. If no answer was given after the time expired, the answer was counted as invalid and not included in this study. After completion of a full round, the participants were able to review their diagnoses and compare them with the correct diagnosis. In the background, the platform stored the selected diagnosis, the current level and round, the answering time, and the screen resolution.

Recruitment, Registration, and Engagement

We used mailing lists, social media posts, and talks at scientific conferences to recruit participants. To compare recruitment strategies, we continuously monitored the number of new registrations and related them to specific recruitment events, if they occurred within 4 days after the event. To analyze registration and dropout, we categorized participants into (1) registered but not verified by email; (2) registered and verified but did not complete any level; (3) registered, verified, played, and completed 1 of the screening levels at least once; and (4) registered, verified, played, and completed all screening levels and the target level at least once. We analyzed engagement with the unbounded retention measure used in game analytics [20] by calculating how many participants returned and played at least one level between 0 and 100 days after their registration. The device types were identified using the free Web analytics software Matomo [21].

Accuracy and Collective Intelligence

We calculated baseline measures of accuracy (ie, correct specific prediction of a disease category, not just malignancy) for each dermatoscopic image and per disease category for single raters. To calculate the measures of accuracy for collective intelligence, we applied bootstrapping (random sampling with replacement), simulating multiple second opinions. We let the size of the collectives range from 3 to 8. Dermatoscopic images for which the number of raters was lower than the size of the virtual collective were excluded. The disease category with most votes (ie, first-past-the-post voting) was regarded as the collective vote per dermatoscopic image; ties were broken at random. This procedure was repeated for answers of nonexperts as well as for the answers with high and low confidence.

We used the answering time as a surrogate measure for the level of confidence for each image. To allow an unbiased comparison, we calculated the mean answering time for every rater individually. Furthermore, if a rater needed more time than the mean individual answering time for the disease category, the level of confidence for a given answer was regarded as low. If, on the other hand, the answering time was lesser than or equal to the individual mean of this specific rater, the level of confidence was regarded as high. The confidence *all* represents all answers, including low- and high-confidence answers.

The primary outcome metric is the mean accuracy, which we defined as the arithmetic mean of accuracies of every image within a rater group. All calculated accuracies per image were compared pairwise with the baseline accuracy of nonexperts. Point estimates for the difference in accuracy, confidence intervals, and *P* values were only calculated for pairwise overlapping images among groups. Sensitivity, specificity, and positive and negative predictive values were used when the number of analyzed classes was binary, as for the case of malignancy. As part of this study, the diagnostic groups *AKIEC*, *BCC*, and *MEL* are regarded as *malignant* and all others (*BKL*, *DF*, *NV*, and *VASC*) as *benign*, ie, a prediction of *BCC* is counted as correct if *MEL* was the ground truth.

As some raters took the target level significantly more often than others, we restricted the number of rounds per rater to 30 to prevent bias. If the rounds were not completed, we included the answers only if more than 50% of cases (ie, 15 dermatoscopic images) per round were rated.

Statistical Methods and Ethics

Classic measures of diagnostic values (sensitivities, specificities, and predictive values) were calculated per rater group and according to standard formulas [22] for detecting a malignant skin lesion, where prediction of any type of malignant disease category (ie, AKIEC, BCC, or MEL) was considered a correct prediction for any malignant image.

Descriptive continuous values are presented as mean with standard deviation; estimates are provided with 95% confidence intervals. We used paired *t* tests for comparing the difference in correct answers for images between single raters and bootstrapped collective intelligence procedures. All *P* values are reported corrected for multiple testing (Bonferroni-Holm [23]) unless otherwise specified, and a 2-sided *P* value <.05 was regarded as statistically significant. Calculations and plotting were performed using R version 3.4.0 [22] and ggplot2 [24].

The study was approved by the ethics review boards of the University of Queensland (protocol number 2017001223) and the Medical University of Vienna (protocol number 1804/2017). During registration, human raters provided written consent to allow analyzing anonymized ratings. A total of 4 participants demanded all their data to be deleted; therefore, their ratings are not included in this study.

Results

Registration, Recruitment, and Engagement

Of the 2497 individuals (1538/2497, 61.59% female) who registered between June 15, 2018, and June 14, 2019, 44.09% (1101/2497) were board-certified dermatologists, 25.55% (638/2497) were dermatology residents, and 16.58% (414/2497)

were general practitioners. In the 365 days, the survey page was visited 21,948 times. The raters came from 5 continents (Africa, n=112; Asia, n=204; Europe, n=1260; Americas, n=594; and Australia/Oceania, n=327).

The raters used mobile phones in 56.80% (13,042/22,961), a desktop computer in 30.80% (7061/22,925), and a tablet in 6.70% (1546/23,074) of visits to the survey page. The mean time spent on the site per visit was 4 min 37 seconds (SD 3 min 2 seconds). Of the 2497 registered raters, 367 (14.69%) dropped out before playing at least one level, 1330 (53.26%) completed the screening levels and started playing the target level, and 1245 (49.85%) completed the target level at least once. The distribution of age, gender, continent of origin, and experience was similar among registered raters who finished the screening tests and played the target level and those who dropped out. Peaks of registrations could be attributed to specific recruitment events. Most participants were recruited from social media (701/2497, 28.07%) or through mailing lists (732/2497, 29.31%). Only 1.96% (49/2497) of the participants were recruited from scientific meetings; the remaining 40.64% (1015/2497) could not be attributed to a specific event. The highest number of participants recruited per day was 676, after a social media campaign. Without any social media marketing, the number of visitors spanned from 15 to 40 visitors per day (at the time of submission). Participants with less than 1 year of experience had the lowest 30-day unbounded retention rate (21.9%), and participants with less than 3 years of experience had the highest 30-day unbounded retention rate (33.7%).

In the target level, we collected 111,019 single ratings, with a mean of 25.2 (SD 18.5) ratings per image. Only the 4216 images with 8 or more ratings were included in this analysis (AKIEC, n=327; BCC, n=514; BKL, n=1099; DF, n=115; MEL, n=1113; NV, n=907; and VASC, n=142). At the nonexpert level, data of 1208 participants, 4216 different images, 4102 rounds, and 101,271 ratings were included. At the expert level, data of 37 participants, 2609 different images, 193 rounds, and 4762 ratings were included.

Collective Human Intelligence

Table 1 shows the mean accuracies achieved by single experts, single nonexperts, and collectives of nonexperts with different group sizes. We did not calculate the accuracies for the collective intelligence of experts as individual images were not seen by a sufficient number of experts.

Table 1. Comparison of mean accuracy of single nonexperts to mean accuracy of different collective sizes and confidence levels. *P* values denote paired t test comparing the number of correct specific diagnoses per overlapping dermatoscopic image.

Experience	Collective size	Confidence[a]	Mean accuracy, %	Mean difference (95% CI)	*P* value
Nonexperts	—[b]	All	58.60	Reference	Reference
Nonexperts	4	All	64.93	6.33 (6.09 to 6.57)	<.001
Nonexperts	8	All	68.51	9.91 (9.52 to 10.29)	<.001
Nonexperts	—	Low	51.90	−6.20 (−6.77 to −5.64)	<.001
Nonexperts	4	Low	56.01	−2.10 (−2.72 to −1.47)	<.001
Nonexperts	8	Low	59.27	1.16 (0.45 to 1.88)	.007
Nonexperts	—	High	61.40	2.77 (2.44 to 3.09)	<.001
Nonexperts	4	High	65.85	7.22 (6.77 to 7.66)	<.001
Nonexperts	8	High	68.40	9.77 (9.21 to 10.32)	<.001
Experts	—	All	68.36	9.43 (8.11 to 10.74)	<.001
Experts	—	Low	55.61	4.67 (2.27 to 7.06)	<.001
Experts	—	High	74.08	11.91 (10.43 to 13.38)	<.001

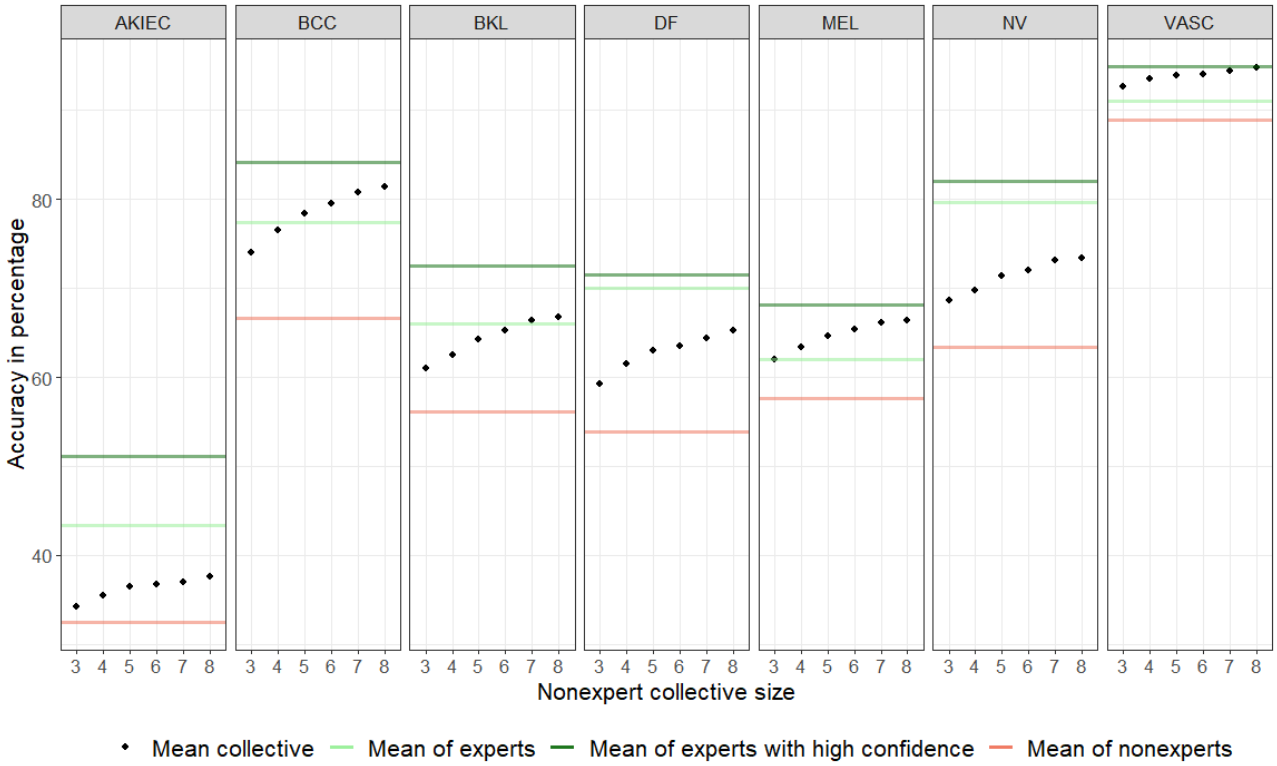
[a]Confidence groups denote whether all answers of raters were measured (All) or only answers given with low or high confidence.

[b]No collective size.

Nonexperts with low confidence had the lowest mean accuracy (51.9%, SD 28.9), whereas confident experts had the highest mean accuracy (74.1%, SD 41.7). A *collective of 8 confident nonexperts* had a similar mean accuracy difference to *single nonexperts* as *single experts* (+9.77 vs +9.43; *P*=.73).

Figure 1 shows the mean sensitivity for each of the 7 disease categories for collectives ranging from 3 to 8 compared with the mean sensitivity of single experts and nonexperts. For all disease categories, the mean sensitivity improved with increasing size of the collective. The mean sensitivity for the disease categories VASC and MEL for collectives of 3 raters was already higher than that for single experts. For BCC and BKL, a collective of 5 and 7, respectively, was needed to surpass the mean accuracy of single experts.

Figure 1. Mean accuracy (dots) per disease category of nonexpert collectives (ranging from 3 to 8) compared with the mean sensitivity of single experts, single experts with high confidence, and single nonexperts. AKIEC: actinic keratosis/intraepithelial carcinoma; BCC: basal cell carcinoma; BKL: benign keratinocytic lesions; DF: dermatofibroma; MEL: melanoma; NV: nevus; VASC: vascular lesions.



The mean time to answer a case was below 5 seconds for both nonexperts (4.7 seconds, SD 4.05) and experts (3.9 seconds, SD 3.47), below 3 seconds in case of high confidence (2.8 seconds, SD 1.75 and 2.3 seconds, SD 1.29, respectively), and above 7 seconds in case of low confidence (7.0 seconds, SD 4.74 and 7.0 seconds, SD 4.22, respectively). The sensitivity for malignant cases increased for both nonexperts and experts in the presence of high confidence compared with low confidence (low vs high nonexperts: low 66.3% vs high 77.6% and experts: low 64.6% vs high 79.4%; see Table 2).

Table 2. Diagnostic values measuring detection of malignant skin lesions for different confidence levels.

Experience	Collective size	Confidence[a]	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive values, % (95% CI)	Negative predictive values, % (95% CI)
Nonexpert	—[b]	All	73.1 (72.7 to 73.5)	77.4 (77.0 to 77.7)	75.1 (74.7 to 75.5)	75.5 (75.1 to 75.9)
Nonexpert	4	All	74.6 (73.9 to 75.2)	78.0 (77.5 to 78.5)	74.5 (73.9 to 75.2)	78.0 (77.5 to 78.6)
Nonexpert	8	All	76.9 (76.3 to 77.5)	80.1 (79.6 to 80.6)	77.0 (76.4 to 77.6)	80.1 (79.5 to 80.6)
Nonexpert	—	Low	66.3 (65.6 to 67.0)	69.7 (69.1 to 70.4)	69.4 (68.7 to 70.1)	66.7 (66.0 to 67.3)
Nonexpert	4	Low	68.7 (68.0 to 69.4)	74.6 (74.0 to 75.2)	70.3 (69.7 to 71.0)	73.1 (72.5 to 73.7)
Nonexpert	8	Low	70.9 (70.3 to 71.6)	76.5 (75.9 to 77.0)	72.5 (71.9 to 73.2)	75.0 (74.4 to 75.6)
Nonexpert	—	High	77.6 (77.1 to 78.0)	81.6 (81.2 to 82.0)	78.7 (78.3 to 79.2)	80.5 (80.1 to 81.0)
Nonexpert	4	High	76.9 (76.3 to 77.5)	77.6 (77.0 to 78.1)	74.8 (74.2 to 75.4)	79.5 (79.0 to 80.0)
Nonexpert	8	High	78.3 (77.7 to 78.8)	79.0 (78.4 to 79.5)	76.3 (75.7 to 76.9)	80.8 (80.2 to 81.3)
Expert	—	All	74.0 (72.2 to 75.8)	85.8 (84.4 to 87.2)	83.1 (81.4 to 84.7)	77.8 (76.2 to 79.4)
Expert	—	Low	64.6 (61.3 to 67.8)	77.4 (74.3 to 80.3)	75.6 (72.2 to 78.7)	66.9 (63.7 to 70.0)
Expert	—	High	79.4 (77.2 to 81.4)	89.7 (88.2 to 91.1)	87.1 (85.2 to 88.9)	83.3 (81.5 to 85.0)

[a]Confidence groups denote whether all answers of raters were measured (All) or only answers given with low or high confidence.

[b]No collective size.

Discussion

Principal Findings

In this study, we showed that collective human intelligence increases the accuracy of nonexperts for the diagnosis of pigmented skin lesions. As collectives, nonexperts reached expert-level accuracies. Although experts were significantly more accurate than nonexperts in general, this difference vanished when average experts were compared with collectives of 8 nonexperts (Table 1). For specific diagnoses, a group of 3 to 8 nonexperts surpassed the sensitivity of the average expert (Figure 1). Potentially, this information could be used for telemedical applications, where a small number of overburdened experts evaluate the majority of referred cases. Small and dynamic groups of physicians, regardless of their expertise, could be used as an alternative to experts. With this strategy, it is possible to recruit raters from a large pool of physicians.

We also found that not all ratings from nonexperts were equally helpful. The ratings given with lower confidence, defined as slow answers in comparison with the mean answer time of a rater, did not increase the accuracy of the collectives of nonexperts. The ratings given with lower confidence even reduced the mean accuracy of small collectives (Table 1). As a consequence, such nonconfident answers should probably be omitted in telemedical applications, which would demand tracking of the time a rater takes to reach a decision. A limitation is that our study rating platform deviated from the real telemedicine platforms as the raters had a maximum of 25 seconds to answer, no option for explanation, and no liability for potential misdiagnoses. Our results with regard to answer time are in line with previous research on a clinical practice colloquially called “if in doubt—cut it out.” Moscarella et al [25] suggested to excise or biopsy lesions for which a specific benign diagnosis cannot be made with confidence. From the data presented herein, the decision boundary on whether there is any *doubt* may be at about 3 seconds. As we also show markedly decreased specificity in low-confidence ratings, a mandatory biopsy in those situations may increase unnecessary interventions; instead, alternative assessments may be sought, such as comparison with other lesions, follow-up imaging, or reflectance confocal microscopy.

Strengths and Limitations

In practice, collective intelligence models, as simulated herein, could be harnessed in different ways to obtain second opinions in difficult cases. Although our method is mainly suitable for store-and-forward approaches, one could also think of real-time simultaneous interaction among readers to possibly further increase accuracy [11]. However, such an approach of swarm intelligence would require participants to be engaged continuously throughout the decision process, evaluating and reevaluating their answer depending on the real-time input of the other participants. Such an interactive swarm could be considered closer to a live discussion with colleagues, whereas the collectives simulated in this study are closer to a classic store-and-forward telemedical approach. Our collectives currently have unclear liability in case of a misdiagnosis: would each member of a collective, the provider of a platform offering collectives, or solely the treating physician be accountable for a misdiagnosis? This conundrum will be equally interesting for computer-aided diagnostics as it is conceivable that one or more members of such collectives could be replaced by a computer algorithm behind the scenes.

Our results are promising with regard to the detection of malignant skin lesions. A collective of 8 confident raters was able to raise single nonexperts’ sensitivity from 73.1% to 78.3% and specificity from 77.4% to 79.9%. Interestingly, although the mean specific accuracy of 8 confident nonexperts was at the level of experts (+0.34% difference), their operating point regarding sensitivity and specificity was more in favor of sensitivity. Therefore, such a nonexpert collective would detect more malignant skin lesions at the cost of more interventions. This, however, may be mitigated by a second line of assessments.

The diagnostic accuracy alone will not be the only consideration in a potential implementation of collective ratings in practice. Although the availability of nonexperts is higher than that of experts, the more the nonexperts involved, the higher the costs and the longer it will take to get a collective vote. With regard to the optimal number of nonexperts, the benefits, such as gain in accuracy for each additional rater, will have to be weighed against these costs. For example, 4 confident nonexperts increase the sensitivity substantially in comparison with a single unconfident nonexpert (from 66.3% to 76.9%), but the additional gain achieved by 8 nonexperts is only marginal (78.3%).

In this study, only dermatoscopic images of pigmented skin lesions were included; however, we estimate that similar improvements are possible with nonpigmented tumors and inflammatory diseases [26], which are more challenging in clinical practice, as shown in previous experiments [27]. A remaining obstacle for application in these areas is a much greater number of possible diagnoses.

We also demonstrated that a high number of raters could be attracted by online marketing and by including elements of gamification. Readers with little experience had a lower unbounded retention rate, which can probably be enhanced by adding additional elements of gamification such as avatars, progress bars (ie, Zeigarnik effect [28]), and better individual metrics or adjusting the learning difficulty level to match the participants’ level.

Conflicts of Interest

None declared.

References

1. Tan E, Levell NJ. Regular clinical dermatoscope use with training improves melanoma diagnosis by dermatologists. Clin Exp Dermatol 2009 Dec;34(8):e876-e878. [doi: 10.1111/j.1365-2230.2009.03629.x] [Medline: 20055853]
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 2;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
3. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol 2019 Jul;20(7):938-947. [doi: 10.1016/S1470-2045(19)30333-X] [Medline: 31201137]
4. Hendrycks D, Gimpel K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. Proc Int Conf Learn Represent 2017:- [FREE Full text]
5. Gyllencreutz JD, Paoli J, Bjellerup M, Bucharbajeva Z, Gonzalez H, Nielsen K, et al. Diagnostic agreement and interobserver concordance with teledermoscopy referrals. J Eur Acad Dermatol Venereol 2017 May;31(5):898-903. [doi: 10.1111/jdv.14147] [Medline: 28150389]
6. Börve A, Gyllencreutz JD, Terstappen K, Backman EJ, Aldenbratt A, Danielsson M, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190 [FREE Full text] [doi: 10.2340/00015555-1906] [Medline: 24923283]
7. Coustasse A, Sarkar R, Abodunde B, Metzger BJ, Slater CM. Use of teledermatology to improve dermatological access in rural areas. Telemed J E Health 2019 Nov;25(11):1022-1032. [doi: 10.1089/tmj.2018.0130] [Medline: 30741608]
8. Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. J Am Acad Dermatol 2011 Jun;64(6):1068-1073. [doi: 10.1016/j.jaad.2010.03.039] [Medline: 21440329]
9. Malone TW, Bernstein MS, editors. Handbook of Collective Intelligence. Cambridge: MIT Press; 2015.
10. Leimeister J. Collective intelligence. Bus Inf Syst Eng 2010;2(4):245-248 [FREE Full text] [doi: 10.1007/s12599-010-0114-8]
11. Rosenberg L, Lungren M, Halabi S, Willcox G, Baltaxe D, Lyons M. Artificial Swarm Intelligence Employed to Amplify Diagnostic Accuracy in Radiology. In: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference. 2018 Presented at: IEMCON; November 1-3, 2018; Vancouver, BC URL: https://doi.org/10.1109/IEMCON.2018.8614883 [doi: 10.1109/IEMCON.2018.8614883]
12. Medical University of Vienna Department of Dermatology. DermaChallenge. URL: https://dermachallenge.meduniwien.ac.at [accessed 2019-07-20]
13. Laravel - The PHP Framework For Web Artisans. URL: https://laravel.com/ [accessed 2019-07-20]
14. React - A JavaScript library for building user interfaces. URL: https://reactjs.org/ [accessed 2019-07-20]
15. Semantic UI React. URL: https://react.semantic-ui.com/ [accessed 2019-07-20]
16. Redux - A Predictable State Container for JS Apps. URL: https://redux.js.org/ [accessed 2019-07-20]
17. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018 Aug 14;5:180161 [FREE Full text] [doi: 10.1038/sdata.2018.161] [Medline: 30106392]
18. Tschandl P. Harvard Dataverse. 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T [accessed 2019-12-11]
19. The International Skin Imaging Collaboration. URL: https://www.isic-archive.com/ [accessed 2019-07-02]
20. Shiu A. Amplitude. 3 Ways To Measure User Retention. URL: https://amplitude.com/blog/2016/08/11/3-ways-measure-user-retention [accessed 2019-07-02]
21. Matomo. URL: https://matomo.org/ [accessed 2019-07-20]
22. R Core Team. R: A language and environment for statistical computing. URL: https://www.R-project.org/ [accessed 2019-07-20]
23. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6(2):65-70 [FREE Full text]
24. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2016.
25. Moscarella E, Lallas A, Longo C, Zalaudek I, Caccavale S, Gambardella A, et al. Performance of the 'if in doubt, cut it out' rule for the management of nodular melanoma. Dermatol Pract Concept 2017 Jul;7(3):1-5 [FREE Full text] [doi: 10.5826/dpc.0703a01] [Medline: 29085713]
26. Errichetti E, Zalaudek I, Kittler H, Apalla Z, Argenziano G, Bakos R, et al. Standardization of dermoscopic terminology and basic dermoscopic parameters to evaluate in general dermatology (non-neoplastic dermatoses): an expert consensus on behalf of the International Dermoscopy Society. Br J Dermatol 2019 May 11:-. [doi: 10.1111/bjd.18125] [Medline: 31077336]
27. Sinz C, Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, et al. Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. J Am Acad Dermatol 2017 Dec;77(6):1100-1109. [doi: 10.1016/j.jaad.2017.07.022] [Medline: 28941871]
28. Lewin K. [Investigations on the psychology of action and affection: III. The memory of completed and uncompleted actions]. Psychol Forsch 1927;9(1):1-85 [FREE Full text] [doi: 10.1007/BF02409755]

Abbreviations

AKIEC: actinic keratosis/intraepithelial carcinoma
BCC: basal cell carcinoma
BKL: benign keratinocytic lesions
DF: dermatofibroma
HAM10000: Human Against Machine with 10000 training images
MEL: melanoma
NV: nevus
VASC: vascular lesions

Edited by G Eysenbach; submitted 23.07.19; peer-reviewed by R Karls, C Longo, L Migliorelli, A Finnane; comments to author 14.09.19; revised version received 05.10.19; accepted 22.10.19; published 24.01.20.

Please cite as:
Rinner C, Kittler H, Rosendahl C, Tschandl P
Analysis of Collective Human Intelligence for Diagnosis of Pigmented Skin Lesions Harnessed by Gamification Via a Web-Based Training Platform: Simulation Reader Study
J Med Internet Res 2020;22(1):e15597
URL: http://www.jmir.org/2020/1/e15597/
doi:10.2196/15597
PMID:32012058

Original Paper

“Asking Too Much?”: Randomized N-of-1 Trial Exploring Patient Preferences and Measurement Reactivity to Frequent Use of Remote Multidimensional Pain Assessments in Children and Young People With Juvenile Idiopathic Arthritis

Rebecca Rachael Lee[1,2], BSc, MSc, PhD; Stephanie Shoop-Worrall[2,3], BSc, MSc, PhD; Amir Rashid[1,2], BSc, MSc, PhD; Wendy Thomson[4], BSc, MSc, PhD; Lis Cordingley[5], BSc, MSc, PhD

[1]National Institute for Health Research Manchester Musculoskeletal Biomedical Research Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom

[2]Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[3]Centre for Health Informatics, The University of Manchester, Manchester, United Kingdom

[4]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[5]Division of Musculoskeletal and Dermatological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Rebecca Rachael Lee, BSc, MSc, PhD
National Institute for Health Research
Central Manchester University Hospitals NHS Foundation Trust
Manchester Academic Health Science Centre
Stopford Building
Oxford Road
Manchester
United Kingdom
Phone: 44 161275 ext 7757
Email: rebecca.lee-4@manchester.ac.uk

Abstract

Background: Remote monitoring of pain using multidimensional mobile health (mHealth) assessment tools is increasingly being adopted in research and care. This assessment method is valuable because it is challenging to capture pain histories, particularly in children and young people in diseases where pain patterns can be complex, such as juvenile idiopathic arthritis (JIA). With the growth of mHealth measures and more frequent assessment, it is important to explore patient preferences for the timing and frequency of administration of such tools and consider whether certain administrative patterns can directly impact on children’s pain experiences.

Objective: This study aimed to explore the feasibility and influence (in terms of objective and subjective measurement reactivity) of several time sampling strategies in remote multidimensional pain reporting.

Methods: An N-of-1 trial was conducted in a subset of children and young people with JIA and their parents recruited to a UK cohort study. Children were allocated to 1 of 4 groups. Each group followed a different schedule of completion of MPT for 8 consecutive weeks. Each schedule included 2 blocks, each comprising 4 different randomized time sampling strategies, with each strategy occurring once within each 4-week block. Children completed MPT according to time sampling strategies: once-a-day, twice-a-day, once-a-week, and as-and-when pain was experienced. Adherence to each strategy was calculated. Participants completed the Patient-Reported Outcomes Measurement Information System Pain Interference Scale at the end of each week to explore objective reactivity. Differences in pain interference scores between time sampling strategies were assessed graphically and using Friedman tests. Children and young people and their parents took part in a semistructured interview about their preferences for different time sampling strategies and to explore subjective reactivity.

Results: A total of 14 children and young people (aged 7-16 years) and their parents participated. Adherence to pain reporting was higher in less intense time sampling strategies (once-a-week=63% [15/24]) compared with more intense time sampling

strategies (twice-a-day=37.8% [127/336]). There were no statistically significant differences in pain interference scores between sampling strategies. Qualitative findings from interviews suggested that children preferred once-a-day (6/14, 43%) and as-and-when pain reporting (6/14, 43%). Creating routine was one of the most important factors for successful reporting, while still ensuring that comprehensive information about recent pain was captured.

Conclusions: Once-a-day pain reporting provides rich contextual information. Although patients were less adherent to this preferred sampling strategy, once-a-day reporting still provides more frequent assessment opportunities compared with other less intense or overburdensome schedules. Important issues for the design of studies and care incorporating momentary assessment techniques were identified. We demonstrate that patient reporting preferences are key to accommodate and are important where data capture quality is key. Our findings support frequent administration of such tools, using daily reporting methods where possible.

(J Med Internet Res 2020;22(1):e14503) doi:10.2196/14503

KEYWORDS

mHealth; pain; pain assessment; juvenile idiopathic arthritis; patient reported outcomes; pediatrics

Introduction

Chronic Pain in Juvenile Idiopathic Arthritis

Chronic pain is defined as pain that is unpleasant and long lasting with sensory, emotional, cognitive, and social components [1,2]. In children and young people with juvenile idiopathic arthritis (JIA), pain is also unpredictable, which hugely contributes to the burden of living with this long-term disease [3-5]. Children and young people conceptualize their JIA in terms of pain experienced [6], suggesting that for some, this is the most salient feature of their illness. However, health care professionals in pediatric rheumatology sometimes neglect to assess pain because of a lack of time and tools available to do so. In addition, professionals may not perceive pain as a priority for their patients because the focus of consultations may be on disease activity and function rather than pain [7,8]. Pain conversations are difficult to have with patients. The use of assessment tools, which remotely and efficiently collect rich pain data for these patients, could help to overcome this clinical problem.

Multidimensional Mobile Health Pain Assessment

In the general field of pediatric pain, multidimensional assessment using electronic mobile health (mHealth) apps is growing widely and quickly [9]. These tools offer several advantages to researchers, clinicians, and patients, compared with unidimensional, paper-based versions. They enable accurate documentation of complex pain data, reporting patterns and pain fluctuations through time- and date-stamped reports [10-12]; demonstrate better adherence and engagement; avoid recall bias through timely monitoring [13]; and allow for wireless data transport, which can be useful for remote management of symptoms [14].

Remote monitoring of pain for children and young people with JIA is valuable; however, there are some specific administrative challenges for this patient group that need to be further explored before these tools can be effectively implemented. Given that the nature of pain qualities can change daily (and within days) for children with JIA [15], pain assessment should be not only multidimensional but also regular and frequent. However, in diary studies of pediatric pain, there is no consensus on the frequency of reporting. Some studies have adhered to once-a-day reporting, and others have implemented 3-times-a-day sampling strategies [16,17]. Studies also often only collect data for short periods (eg, 1 or 2 weeks) [18]. To our knowledge, no research to date has explored patient preferences in the timing and frequency of ecological momentary assessments (EMAs). EMA is a technique used to assess an individual's current experiences, as they occur in real time and real-world settings [19]. In most chronic pain studies using EMA, the time sampling strategies are based on researcher decisions, which have little rationale, and are limited in regard to what is acceptable and feasible for patients [20].

Frequency of Administration

Decisions regarding the frequency of pain assessment administration rarely appear to be evidence based. This is a particular concern for health care professionals managing children and young people with JIA because they fear that more intense, regular pain assessment may lead to over-reporting or overexaggeration of pain or pain-related problems [7]. This phenomenon is called measurement reactivity, which is defined as a change in the variable being measured, because of the nature of the measurement method [21,22]. To our knowledge, this effect has not been explored in electronic multidimensional pediatric pain assessment, but now that real-time data collection techniques are becoming commonplace in pediatrics [18], it is important to ensure that these assessment methods are not detrimental to well-being. To this end, changes in the degree of interference caused by pain may provide a useful indicator of the impact of frequent pain assessment. Therapists involved in the care of those with JIA use pain interference rather than measures, such as simple pain intensity scales, to evaluate the outcome of their interventions [7,23]. This approach fits with a broader concept of pain as a *motivational state* rather than simply a somatic experience [24].

Study Aim

This study aims to investigate patient preferences, feasibility, and influence of several time sampling strategies in remote multidimensional pain reporting. Feasibility studies are used to determine whether an intervention or method is appropriate for real-world use in particular patient groups [25]. In the context of this research study, feasibility referred to the practicability of different self-report schedules, and an N-of-1 trial design

allowed for a comparison of these within the same individuals [26]. We aimed to explore which pain reporting patterns were nonburdensome for children and young people with a complex long-term disease and why from both patients' and parents' perspective (as little qualitative research into patients' reasons for disengaging exists [20]). We also studied the effects of different pain reporting intensities to investigate whether there was any evidence of measurement reactivity in response to using these tools more or less frequently.

Methods

Study Design

This study was a randomized N-of-1 cross-over trial design, which explored the use of 4 different time sampling strategies: once-a-day, twice-a-day, once-a-week, and as-and-when children and young people had pain. These time sampling strategies were chosen based on earlier pilot work in which children and young people with JIA completed My Pain Tracker (MPT; an mHealth multidimensional pain assessment tool, discussed further in the section Materials and Measures) daily for 1 week and discussed how often they thought would be feasible to complete the tool. This study has been reported in accordance with the Consolidated Standards of Reporting Trials extension for reporting N-of-1 trials (Multimedia Appendix 1) [27].

Sample and Recruitment

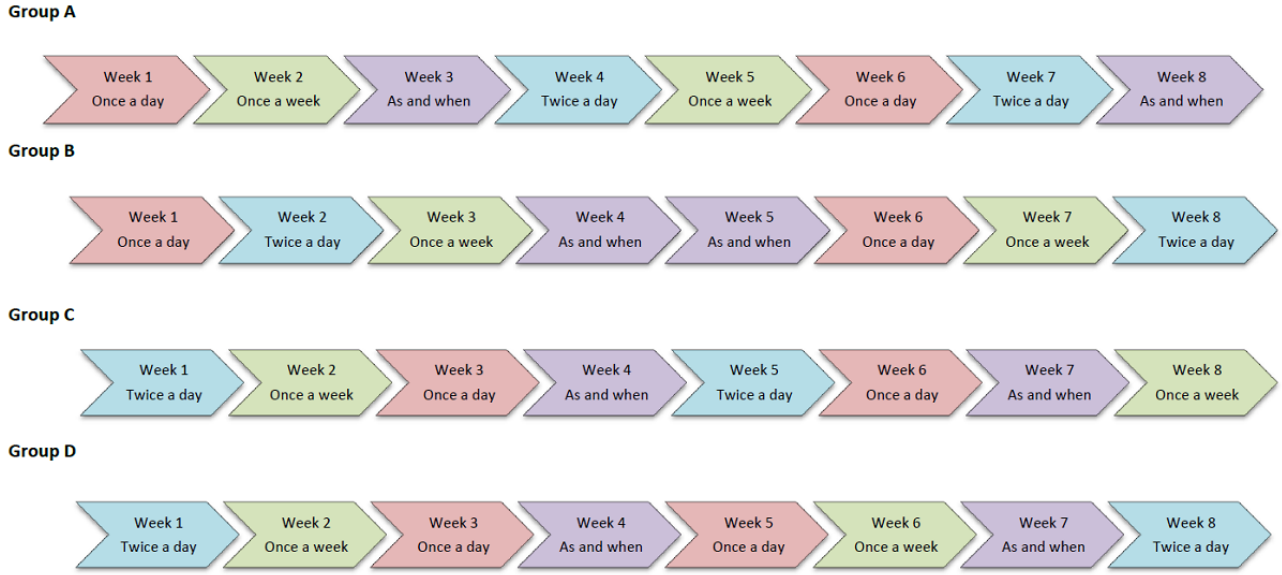
Children and young people and their parents were recruited from a UK prospective inception cohort study of childhood-onset inflammatory arthritis (the Childhood Arthritis Prospective Study [CAPS]). The CAPS study collects data longitudinally from individuals who were diagnosed with inflammatory arthritis (in at least one joint) present for at least 2 weeks under the age of 16 years and who attend 1 of 5 UK pediatric rheumatology centers [28]. Exclusion criteria are septic arthritis, hemarthrosis, and arthritis caused by malignancy/trauma or connective tissue disorders. CAPS began recruitment in 2001 and collects up to 10 years of data following initial presentation to pediatric rheumatology. New children continue to be recruited currently. Written informed consent was provided by proxies for all participants. Assent was also provided by children where appropriate. The study obtained ethical approval as an amendment to CAPS from the UK National Health Service (NHS): Health Research Authority (REC 02/8/104, IRAS 184042).

Eligible children and young people and their parents were identified and contacted by the lead researcher (RRL) to invite them to a participant recruitment event in August 2017 in Manchester, the United Kingdom. Suitability for inclusion in this substudy was based on age (between 5 and 16 years) and English speaking in addition to the CAPS inclusion criteria [28]. Invitation letters and information sheets were sent. Interested children and young people and their parents were encouraged to contact the study team to register and to confirm their attendance at the recruitment event where a presentation about MPTs development was given, consent/assent was taken, and instructional guidance packs were provided. Children and young people and their parents who were interested but unable to attend the event were enrolled in their own homes. These participants were visited by the lead researcher (RRL), who gave a brief demonstration of how MPT worked before enrollment (completion of consent/assent forms and provision of instructional guidance pack for the study).

Randomization

Children and young people were randomized to 1 of 4 groups (group A, B, C, or D), as they presented to the study using a list of random numbers generated before recruitment by the lead researcher (RRL). Each group followed a different pain reporting schedule, which ran for 8 consecutive weeks, divided into 2 blocks of 4 weeks. Each pain reporting schedule included 4 different time sampling strategies (once-a-day, twice-a-day, once-a-week, and as-and-when the pain was experienced), which were randomized to repeat once within each of the 2 blocks. The randomized scheduling was created using block randomization and finalized before children and young people random allocation (see Figure 1). Children and young people were instructed not to complete MPT on the final day of each week (*washout* period) before cross over into another schedule to prevent carryover effects into the next schedule. As participants were not switching between interventions (or on any *active* treatments as part of the study design), only switching between administrative patterns, the anticipation of carryover and lingering effects was minimal. However, this 1-day washout time frame was chosen to balance the risk of any small carryover effects while ensuring that participants did not lose the momentum of using MPT at home. Information about reporting schedules and schedule changes was provided to participants in an instructional guidance pack, which also detailed who to contact if they had problems with MPT or the iPad. All randomization processes were created using Stata version 14.0.

Figure 1. Randomised timing schedules.



Setting

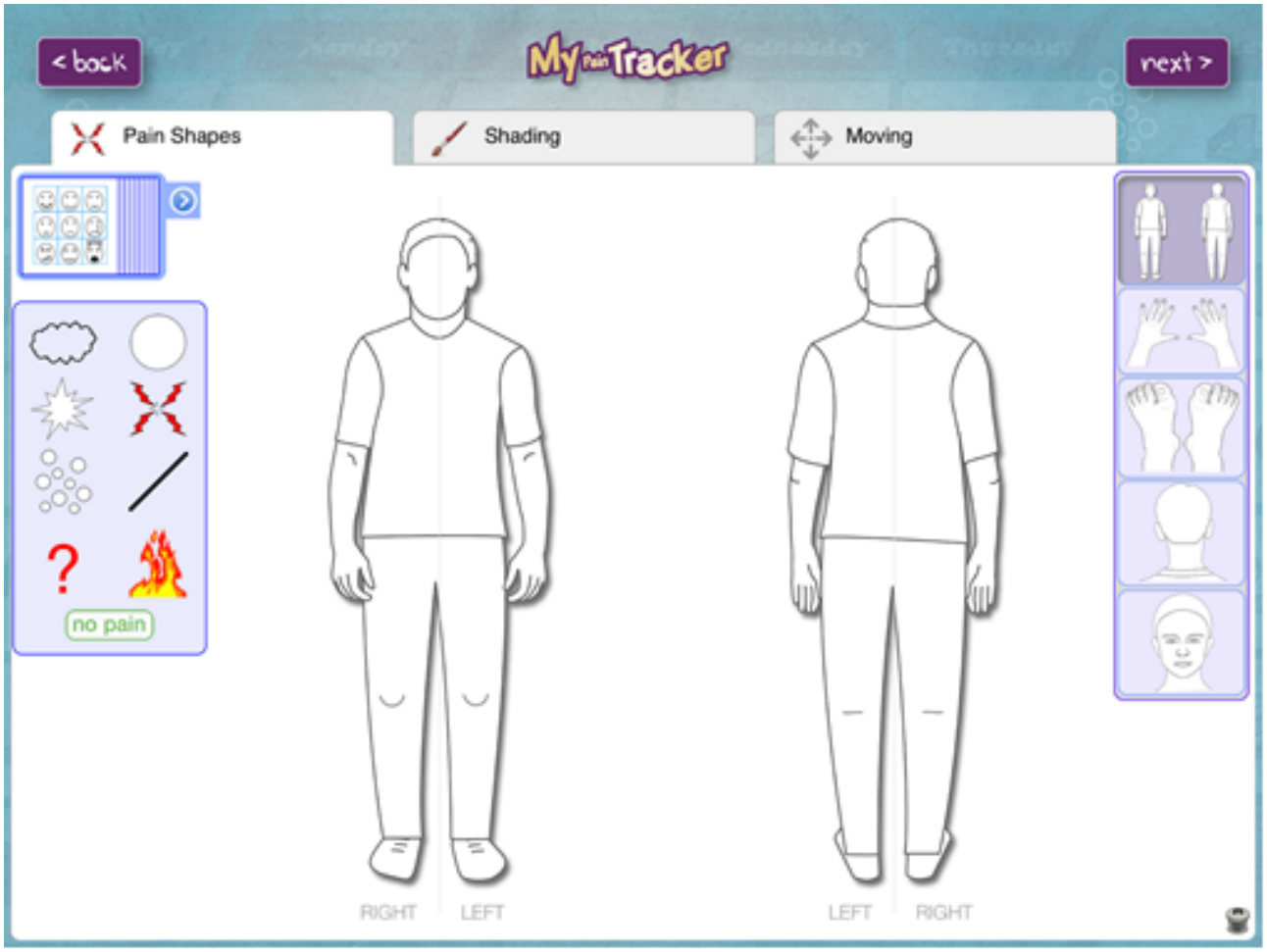
Data collection occurred in participants' own homes throughout the North West of England.

Materials and Measures

Remote Mobile Health Multidimensional Pain Assessment

Children and young people were provided with an iPad with MPT for the duration of the study. MPT is a multidimensional remote monitor of pain for children and young people with JIA. The tool's development is long standing [29-33]. The software underpinning MPT and many of its graphical components first came from a developmentally appropriate computer-aided interview tool developed to facilitate children's communication about somatic symptoms in a mental health context [30]. The tool was then adapted for use in acute postoperative pain, persistent pain, and, more recently, specifically in JIA [29,33,34]. MPT's current format is an iPad app (version 1.6.5), which users manipulate to demonstrate pain experiences. Users of MPT are presented with a body manikin and are able to plot a number of different pain facets on the manikin to represent pain: location, symbols, labels/word descriptors, size (severity), throb rate (intensity), and emotion (see Figure 2 for the main user page of MPT). The app takes approximately 5 min to complete, but children and young people can complete it more or less quickly. After recording their pain using the app, an option is available on the main menu whereby users (including parents) can see their 9 most recent historical pain reports. Participants were contacted by the lead researcher at the end of each week to remind them to change to the new time sampling strategy for the following week.

Figure 2. Main pain reporting page of My Pain Tracker.



Semistructured Interviews

Children and young people and their parents participated in a semistructured telephone interview following 8 weeks of data collection. The interview schedule consisted of questions about which administrative time sampling strategy children and young people and their parents liked the most and why and which was the most appropriate for long-term use of MPT. Children and young people and their parents were asked whether they observed or noted any changes in pain levels in response to the intensity of pain reporting (subjective pain reactivity at the end of the 8-week study period [35]). Telephone interviews lasted between 30 and 45 min and were audio-recorded and transcribed for analysis.

Pain Interference

Pain interference was assessed at the end of each week using the Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Interference Scale-Short Form, which was either completed by children (those aged 8-16 years) or their parents (children younger than 8 years: the PROMIS Parent Proxy Pediatric Pain Interference Scale-Short Form) [36]. This measure was used to assess objective measurement reactivity (whether there were any actual changes in pain-related variables because of different measurement intensities [35]). Both the PROMIS pediatric and parent forms comprised 8 questions relating to the following aspects of pain interference: physical, psychological, and social functioning [36]. To score these measures, questions were scored on a scale of 0 to 4 (never to almost always) and summed for a final score out of 32 (higher scores represent greater pain interference). Scores could be completed in the presence of missing data if at least 50% of items had been answered. In these cases, overall scores were summed and converted to the same scale as if the total number of items had been completed. Previous work has shown that there is strong item agreement between parent- and patient-reported scores on PROMIS measures [37]. Completed questionnaires were returned in prepaid envelopes, as the study was ongoing.

Data Analysis

Statistical Analyses

Adherence was calculated by examining the recorded number of MPT reports, and expected number of MPT entries completed within each time sampling strategy week (apart from as-and-when time sampling strategy, which had no expected number of entries). Quantitative analyses assessed any difference in pain interference scores when pain was recorded at different frequencies. This was primarily assessed graphically, with Friedman tests used to support conclusions drawn. Friedman tests assumed that pain interference did not systematically change week by week over the study duration, and there was no data autocorrelation by the participant. These assumptions were then assessed in secondary analyses.

The relationship between pain interference and time point (weeks in the study) was assessed using scatterplots and Spearman correlations. The reliability of pain interference scores over time within each time sampling strategy was tested using intraclass correlation coefficients (ICCs).

Previous work has suggested that within N-of-1 trials, simple comparisons of means tests outperform more complex methods, such as mixed effects models and meta-analyses, even in the presence of autocorrelation and carryover effects [38]. However, to confirm that autocorrelation at the participant level did not bias the conclusion drawn from the Friedman analysis, a multilevel linear regression tested differences in pain interference between the 4 time-sampling strategies. Strategies were compared against *once-a-day* strategy, and participant number was added as a random effect. All analyses were completed in accordance with an intention-to-treat principle using SPSS version 22 and Stata version 14.0.

Qualitative Analyses

Semistructured telephone interview data were analyzed through deductive semantic thematic analysis. Thematic analysis involves identification of meaningful patterns within data and generates rich, detailed accounts of participant's perspectives [39]. Semantic analysis of data entails that the interpretation of data is largely rooted in the manifest content of participants' accounts [40]. Recurring ideas and topics from children and young people and their parents were identified. These were organized into major and subthemes in NVivo 10. The themes were identified following a deductive approach, meaning predefined themes (about feasibility of time sampling strategies and perceptions of subjective reactive effects) formed the basis of the major themes. Narrative accounts of data phases were written, which were supported by children and young people and their parent quotations.

Results

Participant Characteristics

A total of 373 eligible CAPS participants were contacted to take part in the study, of which 20 potential participants registered. Of 373 participants, 8 children and their parents were able to attend a recruitment event, and a further 6 wished to be enrolled in the study at their homes. Moreover, 6 children and young people and their parents did not respond when contacted again directly by the researcher to organize a time to visit them.

Furthermore, 14 children and young people and their parents took part in the study. Children's and young people's demographics and disease characteristics are presented in Table 1. In addition, 2 participants experienced technical difficulties during the study, which meant that MPT data did not save to iPads. This highlighted issues with the technical feasibility of some of the MPT software and the tool itself (but this was not linked to the feasibility of the time sampling strategies tested). Therefore, MPT data from these 2 participants were excluded from the analysis of adherence. Their data were still included in other statistical analyses, as pain interference questionnaires were returned, and in the qualitative analyses, as semistructured interviews were still conducted. All data were collected between August 2017 and January 2018. The full age range for the participants included in the study was 7 to 16 years.

Table 1. Children and young people's demographics and disease characteristics (N=14).

Characteristics	Values
Age at the study (years), median (IQR)	12.5 (10.0-14.0)
Disease duration at the time of the study (years), median (IQR)	4.3 (2.8-7.0)
Female, n (%)	9 (64)
Subtypes of arthritis, n (%)	
Persistent oligoarthritis	6 (43)
Extended oligoarthritis	1 (7)
RF[a]-negative polyarthritis	5 (36)
RF-positive polyarthritis	1 (7)
Psoriatic arthritis	1 (7)

[a]RF: rheumatoid factor.

Adherence

Overall adherence (n=12) for each time sampling strategy demonstrated that adherence to once-a-week reporting was highest (15/24, 63% possible reports) followed by once-a-day (85/168, 50.6% total possible reports) and twice-a-day (127/336, 37.8% possible reports) reporting. As-and-when reporting ranged from 0 to 7 reports during the 2 weeks this strategy occurred for participants.

Measurement Effects of Pain Monitoring Frequency on Pain Interference

There were no systematic differences in pain interference between children and young people (n=14) in any of the different pain time sampling strategies using MPT (Figure 3; *P*=.77).

There was no correlation between week of participation and pain interference score (r=−0.04; *P*=.68). All time sampling strategies generated high test-retest reliability (all ICC over 0.6; see Table 2). When autocorrelation at the patient level was

accounted for in a multilevel regression model, there were no significant differences in pain interference between time sampling strategies, compared with once-a-day strategy (compared with once-a-day strategy: once-a-week, *P*=.98; twice-a-day, *P*=.59; and as-and-when pain is experienced, *P*=.56).

Figure 3. Boxplots of pain interference score distribution.

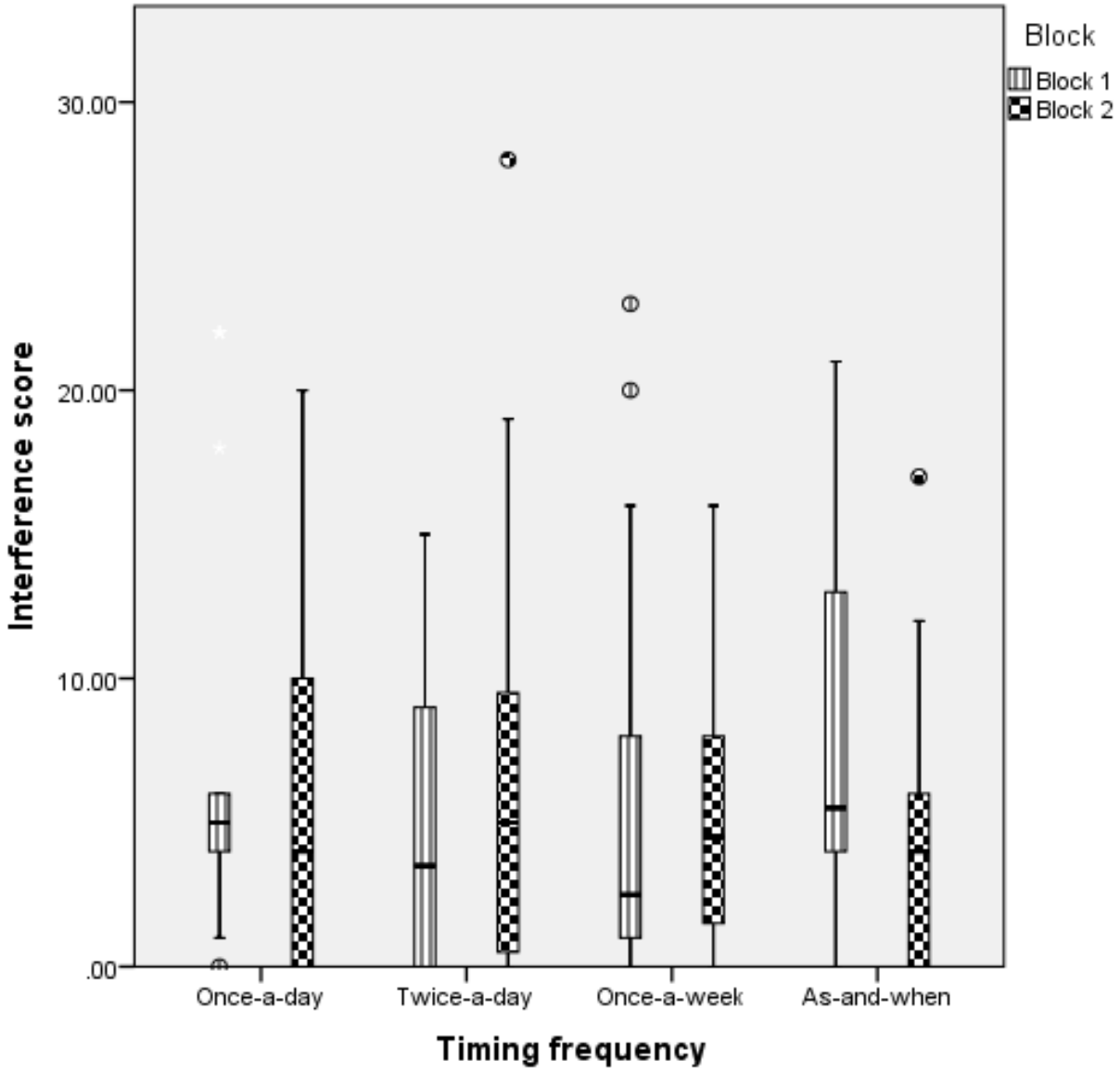


Table 2. Median (IQR) of pain interference scores and correlation coefficients per time sampling strategy.

Time sampling strategy	Median (IQR) (maximum score of 32)			Intraclass correlation coefficient (significance)
	Block 1	Block 2	Blocks 1 and 2 combined	
Once-a-day	5.0 (3.0-12.0)	4.0 (0.0-12.5)	5.0 (3.0-10.0)	0.934
Twice-a-day	3.5 (0.0-9.0)	5.0 (0.0-13.0)	5.0 (0.0-9.0)	0.733
Once-a-week	2.5 (0.8-10.0)	4.5 (0.8-9.0)	5.0 (1.0-10.0)	0.630
As-and-when pain is experienced	5.5 (3.0-13.0)	4.0 (0.0-7.5)	5.0 (2.0-9.0)	0.898

Overview of Qualitative Themes

A total of 2 major themes were deducted from the manifest content of interview data: theme 1: perceived advantages/disadvantages of each time sampling strategy and theme 2: perceived changes in pain experiences during the study (subjective measurement reactivity). Children and young people and their parent preferences for pain reporting strategies are presented in theme 1, and perceptions of subjective measurement reactivity are presented in theme 2, alongside narratives of findings (subthemes). Supporting interview quotations for each theme are presented in Multimedia Appendix 2.

Theme 1: Perceived Advantages/Disadvantages of Each Time Sampling Strategy

An equal number of children and young people preferred once-a-day and as-and-when reporting (6/14, 43%). About 14% (2/14) of children and young people preferred twice-a-day reporting, with only one (1/14, 7%) participant reporting preference for once-a-week reporting. Within parents, 36% (5/13) reported that they most preferred once-a-day reporting, and slightly fewer preferred as-and-when strategy (4/13, 29%). Moreover, 14% (2/13) of parents preferred twice-a-day reporting, with one of these parents suggesting that an even more intensive time sampling strategy would have been appropriate (3 times a day). One parent (1/14, 7%) had no opinion on frequency of administration. Another parent preferred not to speak to the researcher about their preferences for reporting to encourage their child to be independently involved with the study by providing feedback about their use of the app on their own. This parent still provided consent and was present for the young person's enrollment into the study.

Subthemes

Once-a-Day

Reflecting upon, capturing, and storing comprehensive pain information appeared to be valuable to children and young people and meant they could *forget* about their pain until the next day. Children and young people found that they could capture pain variations between days, an advantage that could not be accommodated by less intensive time sampling strategies. Once-a-day reporting appeared to be the easiest to remember to complete because it became routine, whereby other regular activities (such as going to bed) provided a cue. For many of the children and young people, completion when they did not have pain became problematic, and completing the app daily became redundant in these cases. Children and young people who experienced pain daily suggested that MPT should include the option to report on how pain might have changed from morning to evening in this schedule. This seemed to show that although for some children and young people, once-a-day reporting was most feasible, capturing within-day variations in pain was still important.

Parents thought that the shorter recall period made the process of comprehending and reporting pain experiences much easier for children and young people. Some parents believed that longer time between reporting made it more difficult to think about what had happened in pain experiences since children had last reported. The preference for the once-a-day time sampling strategy seemed to be apparent for parents regardless of the level of pain, and as for some parents, it was also useful to know their child was not in pain.

Twice a Day

For some children and young people, twice-a-day reporting was most valuable because it captured within-day variations of pain (eg, the difference between morning and evening pain), which was not useful for those whose pain did not change. For other children and young people, it was difficult to space twice-a-day reporting out, so they did not complete 2 consecutive reports too closely together.

Many parents of children and young people expressed concern for how they would be able to manage this intensive reporting schedule when they were busy. Parents believed that because children were too busy to report, the child would rush the pain report and not show exactly how the pain felt. Owing to the intensity of this time sampling strategy, the information captured may have been less meaningful because less effort was put into completion when pain tracking was too frequent.

Once-a-Week

With this less intensive time sampling strategy, children and young people felt less pressure to fill MPT in, and they could *escape* thinking about pain. However, once-a-week reporting made it difficult to appreciate changes in pain from day to day. The amount of pain information condensed into the 1 weekly report could be problematic, particularly in those with multi-site pain. Children and young people would forget what data they had input during some weeks (how often they had reported and whether they still needed to report at all) or how their pain had been (what types of pain they had, how often, and where the pain was).

As-and-When Pain Is Experienced

The main advantage of as-and-when pain is experienced strategy was reporting flexibility. Some children and young people thought it was useful compared with reporting at times where they might not necessarily have experienced pain. Some children and young people liked that there was not a set amount of times that they had to report pain. The unpredictability of this timing schedule meant that the decision making of when to report was entirely upon the child, which for some was burdensome. For others, the flexibility of the time sampling strategy was problematic because they found it challenging to judge whether they had sufficient pain to report. Some children and young people believed that they should only report pain when it was bad. This was a disadvantage for some because it also meant that this time sampling strategy failed to capture information about when pain was better than usual.

Some children and young people did not like as-and-when reporting because this could be problematic when they were in school. They would forget about pain by the time they were home. For some, they would forget both about the pain they had experienced and to report because they were not prompted to do so. Some parents particularly preferred the as-and-when reporting because they felt that this kind of reporting alerted them to their child being in pain.

Theme 2: Perceived Changes in Pain Experiences During the Study

The majority of children and young people (n=11) and their parents (n=11) did not perceive their pain to have been influenced by assessment frequency. For those who did report subjective reactivity (*children and young people*=3 and *parents*=2), it became apparent that there were 2 kinds of subjective measurement reactivity being referred to: cognitive/emotional and actual physical changes in the level of pain.

Subthemes

Cognitive/Emotional Reactivity

For a small number of children and young people (n=3), intensive pain reporting prompted them to think about pain more, which, in turn, made them more aware of it. These participants talked about how they would try to make a conscious effort to not let pain and thoughts about pain interfere with their day. Some children and young people in the study reported that focusing more on pain led them to notice smaller pains, which otherwise would have gone unnoticed or unreported. Children and young people talked about how they would only report pain that was worse than usual because some were used to a constant level of pain living with a chronic condition. Some parents believed that the bigger focus on pain affected their child's mood and fatigue levels. Parents mentioned that mood and tiredness became worse because children were more aware of the pain (which parents believed their children would rather not think about) with higher frequency reporting. With this increase in the awareness of pain, parents talked about the difficulties of knowing whether to offer their child pain relief. These parents discussed how this difficult decision only seemed to arise when their child was more aware of how pain had felt in the more intense frequency time sampling schedules. It seemed that if children were not prompted to focus on pain for assessment, parents had more confidence in knowing that pain in and of itself was bothering their child.

Physical Pain Reactivity

One parent indicated that their child had experienced a reactive effect on pain severity in response to more intense reporting. In this particular interview, the child themselves did not perceive a change to have occurred. Pain was presumed to worsen by this parent because of the increased attention.

Discussion

Principal Findings

Although mHealth tools are increasingly being used to collect pain data in pediatric chronic pain [9], studies have failed to explore the impact of different time sampling strategies with patients and families [20]. It is important to explore this to develop administrative strategies, which balance reporting burden with the highest quality data collection techniques. Findings from this N-of-1 trial of different time sampling strategies suggest that statistically, there is no objective measurement reactivity with different pain reporting frequencies (in terms of pain interference). Qualitative findings suggest that children and young people and their parents have a preference for once-a-day and as-and-when reporting, but the disadvantages of as-and-when reporting (problematic flexibility and difficulties in remembering to report) far outweigh those cited for once-a-day. For some children and young people, there are perceived changes in emotion and fatigue in response to more intense pain reporting (subjective measurement reactivity). Children and young people demonstrate better adherence to less frequent time sampling strategies (once-a-week and once-a-day) in the short term. However, when reporting more flexibly (as-and-when pain is experienced strategy), some children and young people do not report pain at all, and for others, it can be problematic that *good* pain days are not captured.

Qualitative findings suggest that daily pain reporting is most feasible and preferred for children and young people with JIA, and quantitative data support that frequency of reporting has no impact on pain experience. Once-a-day strategy captured rich pain data, which children and young people valued and often looked back on in their pain report history (documented in MPT). This schedule enabled children and young people to explore patterns in pain (which would be useful for pain discussions with parents and/or health care professionals), and administration was perceived to be easily incorporated into routine. Although an equal proportion of children and young people preferred as-and-when reporting, there were more disadvantages cited for this time sampling strategy overall. For some, the flexibility of reporting was useful, but for others, this aspect of administration was burdensome because there were challenges associated with deciding when and how to report pain and which pains were significant. There were also difficulties associated with reporting pain experienced during school.

In addition to children's and young people's cited disadvantages, there are several methodological issues with event-based reporting, such as as-and-when pain is experienced (participant-initiated reports in response to pain occurrences [19]). In the interpretation of reports, it is impossible to know whether children did not have any pain or whether they did have pain but did not report it for whatever reason. This challenge would make it difficult to compare pain over time within individuals, which would be unfeasible in a clinical/home setting where there is remote data collection involved. Another problem concerns the data that are reported using as-and-when schedules. It is important to understand how and when children and young people define a painful event as occurring, and this would inevitably differ between participants [19]. As highlighted in the findings, some chose to only report unusually bad pain in as-and-when reporting. This schedule is problematic because the open interpretation of when it is necessary to report means that pain, which may be of interest to clinicians and researchers, is not captured.

In this study, patients' preference was for daily pain reporting, whereas data entries were complete during once-a-week time sampling strategy, which may be an indication of the latter being a less demanding task. A strength of our work is that both subjective perspectives and objective indicators of completion were collected; however, the relative merits of each are important to consider. Lower adherence with a more frequent time sampling strategy still provides a more detailed picture (capturing daily variations) compared with complete but less intensive schedules. For example, a richer dataset is collected when a patient misses 3 of 7 daily reports (less adherent) compared with 1/1 weekly report (more adherent). From both research and health care professionals' perspectives, a daily time sampling strategy with an incomplete dataset may appear more challenging to analyze, but it provides better contextual information about the pain experience overall. This advantage was also identified in the participant accounts.

Comparisons With Prior Work

In the adult literature, studies using electronic pain diaries have found no evidence of measurement reactivity (objective or subjective) in participants' data responses [35,41-43]. These papers, however, focused only on reactivity in terms of effects on pain intensity, rather than exploring the relationship between measurement artifacts and other pain-related variables (such as pain interference). Pain-related variables may offer richer contextual information, which is why we chose to explore it in this study. Measurement of pain interference is now recommended as a key outcome in clinical trials [44] and refers to measurement of the extent to which daily activities are interfered with or limited by pain. The assessment of interference is important, especially given that pain intensity does not necessarily correspond with lived experience [45].

It is difficult to disentangle changes in actual experiences (eg, natural fluctuations in pain interference) and whether these would occur regardless of any potential reactive effect, which is an inherent risk of bias in studies such as this [46]. There are ongoing discussions about how to better control for this bias, but to partly address this issue in this study design, the number of measurements was balanced, and the timing of measurement was randomized within individuals and between groups. Furthermore, participants completed *measurements* of reactivity separate to the intervention, which was being explored (MPT), as they were asked to complete an additional measure (PROMIS pain interference scale), which was administered at weekly time points.

In terms of the implications for clinical pain assessment for those with JIA, this study provides some reassurance for health care professionals that fears about pain focusing or pain overreporting may not be justified [7]. The children and young people and their parents in this study reported few differences in their pain experiences during more intense time sampling strategies. Our findings also have implications for pediatric pain research and clinical assessment of pain. Many pediatric pain studies use scales with time reference periods, which encompass substantial periods, such as 2 weeks or a month's worth of pain [47]. This study showed that once-a-week reporting was one of the least appropriate time sampling strategies from patients' and parents' perspectives because children and young people struggle to condense large time frames of information into 1 singular report. This task is inevitably more difficult in even wider time reference periods.

Strengths and Limitations

The substantial data collection period of this study should be considered a strength, as usually studies drawing on momentary assessment methodology in children and young people with chronic pain have narrow data collection periods (usually between 1, 2, or 3 weeks) [11,17,48]. A further strength of the study was the considerable sample size we recruited for the N-of-1 cross-over trials, which can be a burdensome study design for those involved. Very few N-of-1 trials have been conducted with children and where conducted in adults, it is not unusual to combine trial data from samples of less than 10 participants [26,49]. In N-of-1 trials, the number of data points per participant is considered to be more important than the total sample size [50].

A limitation of data collection is that participants were asked to verbally comment at the end of the study on whether they noticed changes in pain levels in response to the intensity of pain reporting. These subjective perceptions about measurement reactivity were thus based on their memory rather than objective assessment.

A limitation of this specific study design is that we did not explore the impact of random sampling protocols in children and young people (where randomly programmed iPad alerts would have prompted individuals when to report pain). This is the third main sampling type in real-time data collection techniques (in addition to time driven [based on preset schedules, such as once-a-day] and event-triggered [such as as-and-when schedule]). Exploring the feasibility of random scheduling would have allowed for a more complete picture of children's experiences of using a wider range of momentary assessment techniques. A limit of the study findings is that N-of-1 trial findings are applicable to the patients the trial are conducted with and not for identifying population-level conclusions [51]. The findings of this intensive trial design possibly reflect the attitudes and impact of pain assessment frequency in some children and young people with JIA (we included participants of different ages and of different subtypes), but other patients may have different pain reporting needs that need to be investigated.

Future Research

Future research should aim to explore the experience of perceived subjective measurement reactivity and the quality of these perceptions/changes in more detail. Several children and young people in our study reported some changes in mood and fatigue in their perceived response to more intense reporting schedules, which aligns with concerns that attention to pain leads to overexaggeration of symptoms expressed by health care professional's perspectives [7]. Given these concerns, it would be interesting to explore health care professionals' perspectives on these approaches. Further research should identify those patients for whom more frequent assessment of pain might be problematic and particularly emotionally or cognitively demanding. From this, we should develop phenotypes of these individuals to ensure appropriate pain data collection with minimal harm in pediatric pain studies and clinical care using these complex, multidimensional pain assessment tools.

Conclusions

In conclusion, our study highlights that daily reporting of pain using mHealth multidimensional assessments is most feasible in terms of patient preference and adherence in long-term data collection with children and young people with JIA. There was no evidence to support that any timing schedule had an objective impact on pain interference, although there were some perceived changes in mood and fatigue in more intense reporting schedules for some participants. These findings are important for the development of administrative guidelines for remote pain monitoring tools, which accommodate momentary assessment

techniques. Overall, our findings support the use of mHealth multidimensional pain assessment tools regularly and frequently to better capture pain patterns in children and young people with JIA.

Acknowledgments

This work was supported by the National Institute for Health Research Biomedical Research Centre Funding Scheme and the National Institute for Health Research Manchester Biomedical Research Centre. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. Aspects of the work were also supported by grants from Versus Arthritis, the United Kingdom. These include funding for the Centre for Epidemiology Versus Arthritis (grant reference number: 20380) and the Centre for Genetics and Genomics Versus Arthritis (grant reference number: 20385).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Standards of Reporting Trials extension for reporting N-of-1 trials checklist.
[DOCX File , 21 KB - jmir_v22i1e14503_app1.docx]

Multimedia Appendix 2
Supporting interview quotations for each identified theme.
[DOCX File , 29 KB - jmir_v22i1e14503_app2.docx]

References

1. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms. Prepared by the International Association for the Study of Pain, Subcommittee on Taxonomy. Pain Suppl 1986;3:S1-226. [Medline: 3461421]
2. Williams AC, Craig KD. Updating the definition of pain. Pain 2016 Nov;157(11):2420-2423. [doi: 10.1097/j.pain.0000000000000613] [Medline: 27200490]
3. Clinch J, Eccleston C. Chronic musculoskeletal pain in children: assessment and management. Rheumatology (Oxford) 2009 May;48(5):466-474. [doi: 10.1093/rheumatology/kep001] [Medline: 19202161]
4. Schanberg LE, Anthony KK, Gil KM, Maurin EC. Daily pain and symptoms in children with polyarticular arthritis. Arthritis Rheum 2003 May;48(5):1390-1397 [FREE Full text] [doi: 10.1002/art.10986] [Medline: 12746912]
5. Rashid A, Cordingley L, Carrasco R, Foster HE, Baildam EM, Chieng A, et al. Patterns of pain over time among children with juvenile idiopathic arthritis. Arch Dis Child 2018 May;103(5):437-443 [FREE Full text] [doi: 10.1136/archdischild-2017-313337] [Medline: 29175824]
6. Ghio D, Thomson W, Calam R, Ulph F, Baildam EM, Hyrich K, et al. The prioritization of symptom beliefs over illness beliefs: the development and validation of the Pain Perception Questionnaire for Young People. Br J Health Psychol 2018 Feb;23(1):68-87 [FREE Full text] [doi: 10.1111/bjhp.12275] [Medline: 28990337]
7. Lee RR, Rashid A, Thomson W, Cordingley L. 'Reluctant to assess pain': a qualitative study of healthcare professionals' beliefs about the role of pain in Juvenile Idiopathic Arthritis. Arthritis Care Res (Hoboken) 2019 Jan 10:-. [doi: 10.1002/acr.23827] [Medline: 30629337]
8. Fitzcharles M, Almahrezi A, Shir Y. Pain: understanding and challenges for the rheumatologist. Arthritis Rheum 2005 Dec;52(12):3685-3692 [FREE Full text] [doi: 10.1002/art.21435] [Medline: 16329076]
9. 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]
10. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Feb;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005] [Medline: 14736583]
11. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 8;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
12. Shiffman S. Designing protocols for ecological momentary assessment. In: Stone A, Shiffman S, Atienza A, Nebeling L, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research. New York: Oxford University Press; 2007:27-46.
13. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: a smartphone app to assess pain intensity. J Pain 2014 Oct;15(10):1001-1007. [doi: 10.1016/j.jpain.2014.04.009] [Medline: 24854065]

14. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
15. Stinson JN, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, Dupuis A, et al. Construct validity of a multidimensional electronic pain diary for adolescents with arthritis. Pain 2008 Jun;136(3):281-292. [doi: 10.1016/j.pain.2007.07.002] [Medline: 17723279]
16. Lewandowski AS, Palermo TM, Kirchner HL, Drotar D. Comparing diary and retrospective reports of pain and activity restriction in children and adolescents with chronic pain conditions. Clin J Pain 2009 May;25(4):299-306 [FREE Full text] [doi: 10.1097/AJP.0b013e3181965578] [Medline: 19590478]
17. Stinson JN, Jibb LA, Lalloo C, Feldman BM, McGrath PJ, Petroz GC, et al. Comparison of average weekly pain using recalled paper and momentary assessment electronic diary reports in children with arthritis. Clin J Pain 2014 Dec;30(12):1044-1050. [doi: 10.1097/AJP.0000000000000072] [Medline: 24535055]
18. Stinson JN. Improving the assessment of pediatric chronic pain: harnessing the potential of electronic diaries. Pain Res Manag 2009;14(1):59-64 [FREE Full text] [doi: 10.1155/2009/915302] [Medline: 19262918]
19. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
20. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
21. Campbell DT. Factors relevant to the validity of experiments in social settings. Psychol Bull 1957 Jul;54(4):297-312. [doi: 10.1037/h0040950] [Medline: 13465924]
22. Haynes SN, O'Brien WH. Principles and Practice of Behavioral Assessment. New York: Gardner Press; 1978.
23. Lee RR, Rashid A, Ghio D, Thomson W, Cordingley L. 'Seeing Pain Differently': a qualitative investigation into the differences and similarities of pain and rheumatology specialists' interpretation of multidimensional mobile health pain data from children and young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2019 Jul 2;7(7):e12952 [FREE Full text] [doi: 10.2196/12952] [Medline: 31267979]
24. Wall PD. On the relation of injury to pain. The John J. Bonica lecture. Pain 1979 Jun;6(3):253-264. [doi: 10.1016/0304-3959(79)90047-2] [Medline: 460933]
25. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
26. 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]
27. Shamseer L, Sampson M, Bukutu C, Schmid CH, Nikles J, Tate R, CENT Group. CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. Br Med J 2015 May 14;350:h1793 [FREE Full text] [doi: 10.1136/bmj.h1793] [Medline: 25976162]
28. Hyrich KL, Lal SD, Foster HE, Thornton J, Adib N, Baildam E, et al. Disease activity and disability in children with juvenile idiopathic arthritis one year following presentation to paediatric rheumatology. Results from the Childhood Arthritis Prospective Study. Rheumatology (Oxford) 2010 Jan;49(1):116-122 [FREE Full text] [doi: 10.1093/rheumatology/kep352] [Medline: 19926670]
29. Watson S, Calam R, Jimmieson P. Can computers help in assessing children's postoperative pain? Initial validation of a computer-assisted interview. Eur J Anaesthesiol 2002 Jul;19(7):510-516. [doi: 10.1017/s0265021502000832] [Medline: 12113614]
30. Calam RM, Jimmieson P, Cox AD, Glasgow DV, Larsen SG. Can computer-based assessment help us understand children's pain? Eur J Anaesthesiol 2000 May;17(5):284-288. [doi: 10.1046/j.1365-2346.2000.00655.x] [Medline: 10926067]
31. Tapping P, Rashid A, Thomson W, Ghio D, Cordingley L, Calam R, et al. OP0157-HPR 'This Feeling!': can a new iPad app help children with juvenile idiopathic arthritis communicate their pain experiences? Feasibility, usability and acceptability. Ann Rheum Dis 2015;74(Suppl 2):128.2-12128. [doi: 10.1136/annrheumdis-2015-eular.1141]
32. Tapping P, Calam R, Cordingley L, Thomson W, Rashid A, Chieng A. 'This Feeling': Development and Validity Testing of a Pain Assessment App for Juvenile Idiopathic Arthritis. In: Proceedings of the 10th International Symposium on Pediatric Pain. 2015 Presented at: ISPP'15; May 31 - June 4, 2015; Seattle, Washington, USA.
33. Twynholm J. Computer aided assessment of childrens recurrent pain (Thesis submitted). Manchester: University of Manchester; 2009.
34. Tapping P. Research Gate. Manchester: University of Manchester; 2015. Multidimensional Pain Communication in Children and Adolescents. URL: https://www.researchgate.net/publication/306029559_Multidimensional_pain_communication_in_children_and_adolescents [accessed 2019-12-04]
35. Aaron LA, Turner JA, Mancl L, Brister H, Sawchuk CN. Electronic diary assessment of pain-related variables: is reactivity a problem? J Pain 2005 Feb;6(2):107-115. [doi: 10.1016/j.jpain.2004.11.003] [Medline: 15694877]
36. Varni JW, Stucky B, Thissen D, Dewitt EM, Irwin D, Lai J, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. J Pain 2010 Nov;11(11):1109-1119 [FREE Full text] [doi: 10.1016/j.jpain.2010.02.005] [Medline: 20627819]

37. Varni JW, Thissen D, Stucky BD, Liu Y, Magnus B, Quinn H, et al. PROMIS Parent Proxy Report Scales for children ages 5-7 years: an item response theory analysis of differential item functioning across age groups. Qual Life Res 2014 Feb;23(1):349-361 [FREE Full text] [doi: 10.1007/s11136-013-0439-0] [Medline: 23740167]
38. Chen X, Chen P. A comparison of four methods for the analysis of N-of-1 trials. PLoS One 2014;9(2):e87752 [FREE Full text] [doi: 10.1371/journal.pone.0087752] [Medline: 24503561]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Joffe H. Thematic analysis. In: Harper D, Thompson A, editors. Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners. Chichester: Wiley-Blackwell; 2012:209-223.
41. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
42. Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PF, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain 2000 Feb;84(2-3):181-192. [doi: 10.1016/s0304-3959(99)00206-7] [Medline: 10666523]
43. Cruise CE, Broderick J, Porter L, Kaell A, Stone AA. Reactive effects of diary self-assessment in chronic pain patients. Pain 1996 Oct;67(2-3):253-258. [doi: 10.1016/0304-3959(96)03125-9] [Medline: 8951918]
44. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
45. Hoftun GB, Romundstad PR, Zwart J, Rygg M. Chronic idiopathic pain in adolescence--high prevalence and disability: the young HUNT Study 2008. Pain 2011 Oct;152(10):2259-2266. [doi: 10.1016/j.pain.2011.05.007] [Medline: 21683528]
46. Miles LM, Elbourne D, Farmer A, Gulliford M, Locock L, McCambridge J, et al. Bias due to MEasurement Reactions In Trials to improve health (MERIT): protocol for research to develop MRC guidance. Trials 2018 Nov 26;19(1):653 [FREE Full text] [doi: 10.1186/s13063-018-3017-5] [Medline: 30477551]
47. Lee RR, Rashid A, Ghio D, Thomson W, Cordingley L. Chronic pain assessments in children and adolescents: a systematic literature review of the selection, administration, interpretation, and reporting of unidimensional pain intensity scales. Pain Res Manag 2017;2017:7603758 [FREE Full text] [doi: 10.1155/2017/7603758] [Medline: 28912638]
48. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
49. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. PLoS One 2016;11(12):e0167167 [FREE Full text] [doi: 10.1371/journal.pone.0167167] [Medline: 27907056]
50. McDonald S, Quinn F, Vieira R, O'Brien N, White M, Johnston DW, et al. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. Health Psychol Rev 2017 Dec;11(4):307-323. [doi: 10.1080/17437199.2017.1316672] [Medline: 28406349]
51. Vohra S, Punja S. N-of-1 trials: individualized medication effectiveness tests. Virtual Mentor 2013 Jan 1;15(1):42-45 [FREE Full text] [doi: 10.1001/virtualmentor.2013.15.1.stas2-1301] [Medline: 23356806]

Abbreviations

CAPS: Childhood Arthritis Prospective Study
EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
JIA: juvenile idiopathic arthritis
mHealth: mobile health
MPT: My Pain Tracker
NHS: National Health Service
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by G Eysenbach; submitted 07.05.19; peer-reviewed by J Stinson, L König; comments to author 09.07.19; revised version received 28.08.19; accepted 02.10.19; published 30.01.20.

Please cite as:
Lee RR, Shoop-Worrall S, Rashid A, Thomson W, Cordingley L
"Asking Too Much?": Randomized N-of-1 Trial Exploring Patient Preferences and Measurement Reactivity to Frequent Use of Remote Multidimensional Pain Assessments in Children and Young People With Juvenile Idiopathic Arthritis
J Med Internet Res 2020;22(1):e14503
URL: http://www.jmir.org/2020/1/e14503/
doi:10.2196/14503
PMID:32012051

Original Paper

Why Employees (Still) Click on Phishing Links: Investigation in Hospitals

Mohammad S Jalali[1,2], PhD; Maike Bruckes[3], PhD; Daniel Westmattelmann[3], PhD; Gerhard Schewe[3], PhD

[1]Massachusetts General Hospital Institute for Technology Assessment, Harvard Medical School, Boston, MA, United States

[2]Massachusetts Institute of Technology Sloan School of Management, Cambridge, MA, United States

[3]Center for Management, University of Muenster, Muenster, Germany

Corresponding Author:
Maike Bruckes, PhD
Center for Management
University of Muenster
Universitaetsstraße 14-16
Muenster
Germany
Phone: 49 2518323539
Email: maike.bruckes@wiwi.uni-muenster.de

Abstract

Background: Hospitals have been one of the major targets for phishing attacks. Despite efforts to improve information security compliance, hospitals still significantly suffer from such attacks, impacting the quality of care and the safety of patients.

Objective: This study aimed to investigate why hospital employees decide to click on phishing emails by analyzing actual clicking data.

Methods: We first gauged the factors that influence clicking behavior using the theory of planned behavior (TPB) and integrating trust theories. We then conducted a survey in hospitals and used structural equation modeling to investigate the components of compliance intention. We matched employees' survey results with their actual clicking data from phishing campaigns.

Results: Our analysis (N=397) reveals that TPB factors (attitude, subjective norms, and perceived behavioral control), as well as collective felt trust and trust in information security technology, are positively related to compliance intention. However, compliance intention is not significantly related to compliance behavior. Only the level of employees' workload is positively associated with the likelihood of employees clicking on a phishing link.

Conclusions: This is one of the few studies in information security and decision making that observed compliance behavior by analyzing clicking data rather than using self-reported data. We show that, in the context of phishing emails, intention and compliance might not be as strongly linked as previously assumed; hence, hospitals must remain vigilant with vulnerabilities that cannot be easily managed. Importantly, given the significant association between workload and noncompliance behavior (ie, clicking on phishing links), hospitals should better manage employees' workload to increase information security. Our findings can help health care organizations augment employees' compliance with their cybersecurity policies and reduce the likelihood of clicking on phishing links.

(J Med Internet Res 2020;22(1):e16775) doi:10.2196/16775

KEYWORDS

information security management; phishing emails; compliance; trust; theory of planned behavior

Introduction

Background

The digitalization of health records is vastly transforming the health care industry, establishing enhanced treatment results and medical care experiences. By providing and sharing information, digital health care information systems (IS) are beneficial in various ways: they result in less chance for human error, continuous and autonomous monitoring of the patient, and more efficiency [1]. However, the increasingly complex digital systems have also resulted in major security challenges. Health care organizations are especially vulnerable to information security threats, as data breaches can have direct and severe consequences on patients' lives [2-4]. Attacks against

hospitals have been increasing in both number and level of sophistication [5].

Cybersecurity pertains to protecting internet networks and their confidential information from unwanted invasions and accidental leaks [6]. In information security management, humans are the weakest link—any employee who violates information security policies (ISPs) makes their organization vulnerable to a cybersecurity attack [7,8]. Discovering "why" employees fail to comply with ISP is critical in protecting an organization's information.

Phishing emails demonstrate this issue. Phishing is the practice of sending emails claiming a false identity to induce individuals to reveal information. These fraudulent emails are tailored to access information systems by targeting those with access to the system. Phishing poses a major cybersecurity risk for 2 reasons: (1) employees usually have detailed knowledge about IS within the organization and access the data frequently during their work and (2) even 1 innocent click could expose the organization to a network of hackers nearly impossible to trace [9-12]. A recent study analyzed phishing campaigns in health care organizations and found that, on average, as much as 14.2% of these phishing emails were clicked on by employees [5]. Organizations have taken steps to address this problem by providing training programs to educate and increase cybersecurity awareness, but these efforts remain insufficient. In fact, research shows that mandatory training programs did not make a large difference on reducing clicking rates on phishing links [13]. Recent evidence indicates that approximately 70% of hospitals fail to establish or uphold sufficient privacy and security measures [14].

To investigate employee's compliance with ISP, previous research has often focused on cognitive beliefs based on the theory of planned behavior (TPB) [8,9]. TPB has often been validated and is the most commonly used theory to measure different antecedents to ISP compliance [15-18]. However, previous studies have not adequately investigated the components of these cognitive beliefs. One such component is trust. Trust influences how individuals assess cost-benefit considerations and make decisions, and ultimately their behavior [19,20]. Trust has been investigated from a broad range of research directions and has evolved to a widely accepted and established concept [21-24].

Particularly in regard to phishing attacks, 2 major questions remain unanswered: (1) What is the role of trust in predicting employees' compliance intention? and (2) To what extent does compliance *intention* correspond to compliance *behavior*? To address these questions, we drew on the TPB and investigated factors that motivate compliance with information security guidelines. We conducted a survey and used data from phishing campaigns to highlight relationships among employees' attitudes and beliefs and their actual compliance behavior.

The study consists of 2 steps: First, as a part of phishing tests, employees of hospitals received a faux phishing email. Second, about 6 weeks apart, all individuals (clickers and nonclickers) answered a survey that examined their attitudes and positions toward cybersecurity policy. As we were comparing individuals' qualitative answers in the survey against their own clicking data, we were able to observe and compare their compliance intention with their actual behavior.

This paper is organized as follows: We first present the theoretical background and the research model. Next, we present our research methods, including the structure of the phishing ploy and the survey. Finally, we present our data analysis, results, and discussion.

Theory of Planned Behavior

TPB has emerged as one of the most influential frameworks for the explanation of human behavior [25,26]. The TPB explains that attitudes, subjective norms, and perception of behavioral controls (see Ajzen [27] for the definitions of these elements) form an individual's intention to perform a certain behavior—intention is a direct antecedent of the actual behavior. A positive attitude toward ISP is assumed to predict compliance intention. Bulgurcu et al [9] focus on the link between employees' attitudes toward compliance and their intention to comply and found a positive relationship. Similarly, Ifinedo [28] investigated ISP compliance of managers and IS professionals. These studies concluded that attitude toward compliance, subjective norms, and response efficacy positively influence employees' general ISP compliance intentions. Although these findings all show that TPB is generally suitable for predicting intention in information security research, the specific context (ie, phishing) is a major influence on the behavioral intention—as the context might substantially influence the outcome. Thus, we build on previous research by proposing that TPB variables are associated with employees' intention to comply specifically with ISP:

H1a: Attitudes toward ISP is positively related to the intention to comply.

H1b: Subjective norm is positively related to the intention to comply.

H1c: Perceived behavioral controls are positively related to the intention to comply.

H2: The intention to comply is positively related to compliance behavior.

Collective Felt Trust

A second factor we believe influences compliance is collective felt trust. In their review, Mayer et al [29] suggest that trust influences employees' behavior in the sense that it affects risk-taking in relationships and impacts processes and outcomes in an organization. Trust is defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" [24].

Trust has previously been shown to influence attitudes in the TPB. Pavlou et al [30] investigated whether trust is relevant for the attitude toward a certain product. They found that trust in the person providing a product had a significant effect on attitude toward the product. Management of an organization is responsible for providing a work environment within the company that enables employees to focus on their tasks. Moreover, trust has been shown to impact organizational support and commitment [31,32] and organizational citizenship behavior

[33,34]. Meta-analytic evidence has shown that by trusting the management, employees feel more committed to their company and will be more willing to follow organizational policies [35]. Several studies report a positive relationship between trust and compliance [36-39]. In addition to this relationship, Deutsch Salamon and Robinson [40] found that felt trust increased employees' responsibility norms and subsequently their performance. We assume that this effect holds in this context too: Employees that feel trusted by their management consider their behavior more closely to not violate the trust they are being given.

On the basis of these considerations, we argue that collective felt trust influences employees' attitudes toward ISP and their perceived subjective norms. Thus, we propose:

H3a: Collective felt trust is positively related to the attitudes toward ISP.

H3b: Collective felt trust is positively related to subjective norms.

Trust in Technology

Although trust has often been researched on the interpersonal level, recent developments show that trust in technology is equally important [41,42]. Trust in technology has been shown to increase the acceptance and intention to use technologies [41,43], such as information and technology (IT) artifacts [44] and cloud technologies [45].

When individuals find themselves in risky situations in which they have to depend on technologies, trust in technology becomes essential [46]. Individuals are sensitive to the functioning of that specific technology—similar to trust in people, trust in technology is formed based on the perception of the attributes of technology. Lankton et al [22] suggest differentiating among perceptions of functionality, helpfulness, and reliability as factors affecting trust in technology. In the context of information security, the helpfulness (eg, of an antivirus) is rather limited, although functionality and reliability are highly relevant. High trust in technology will enhance the level of confidence in facing cyber threats. We therefore assume that:

H4: Trust in technology, consisting of (1) reliability and (2) functionality, is positively related to perceived behavioral control.

In situations where individuals perceive a higher threat of cyberattacks, the attention toward potential harms might rise. Johnston et al [11] discuss the influence of fear appeals in IS security and argue that the more severe or susceptible a threat is perceived to be, the fewer individuals rely on the ability of the cybersecurity software. Thus, the higher the perceived risk, the more individuals are expected to pay attention to situations where the software did not adequately eliminate the threat, that is, phishing emails. We therefore propose that:

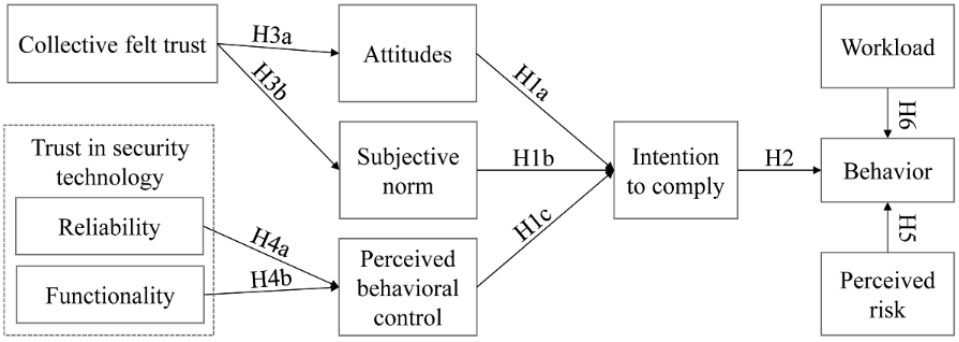
H5: Perceived risk is positively related to the compliance behavior.

On the other side, employees that face a high workload might not be able to execute cognitive considerations to decide to follow ISP. Employees might even use their high workload as an excuse for violating ISP [12]. In situations where high workload stops employees from paying attention to details of an email, whether intentionally or accidentally, the likelihood of opening a potentially dangerous email might increase. Thus, we propose that:

H6: High workload is negatively related to the compliance behavior.

Putting these hypotheses together, Figure 1 presents our proposed research model.

Figure 1. Proposed research model. H: hypothesis.



Methods

Data and Procedure

Data were collected in 2 steps. In the first step, a professional cybersecurity company sent out phishing emails to employees in 3 networks of hospitals in the eastern United States. The phishing campaigns were designed to resemble real phishing emails so that participants could not know that they are being tested and would behave as if they received a real phishing email. All phishing emails contained a hyperlink. Collected data included the identity of individuals receiving the email and whether they clicked on the link or not. This information was then provided only to the respective hospital.

For the second step, we developed a Web-based survey instrument. To compare the results of clickers and nonclickers, we created 2 different survey links based on the same questionnaire. The key constructs with the underlying items are listed in Table 1.

Table 1. Survey items.

Construct and items	Loadings	Cronbach alpha
Attitudes toward information security policy		.86
I believe it is beneficial for our organization to establish clear information security policies, practices, and technologies.[a]	0.891	
I believe it is useful for our organization to enforce its information security policies, practices, and technologies.[a]	0.756	
I believe it is a good idea for our organization to establish clear information security policies, practices, and technologies.[a]	0.884	
Subjective norm		.93
People who influenced my behavior would think that I should follow the policies and procedures and use the cybersecurity technologies.[a]	0.844	
People whose opinions are important to me would think that I should follow the policies and procedures and use the cybersecurity technologies.[a]	0.955	
People whom I respect would think that I should follow the policies and procedures and use the cybersecurity technologies.[a]	0.952	
Perceived behavioral control		.79
I am able to follow the cybersecurity policies and procedures and technologies (eg, antivirus, or other products).[a]	0.665	
I have the resources and knowledge to follow the policies and procedures and use the cybersecurity technologies.[a]	0.917	
I have adequate training to follow the policies and procedures and use cybersecurity technologies.[a]	0.850	
Intention to comply		1
I intend to follow the information security policies and practices at work.[c]	1	
Collective felt trust		.77
Management lets me have an impact on issues they find important.[a]	Dropped	
Management does not feel the need to *keep an eye* on me.[a]	0.773	
Management would be comfortable assigning me a critical task, even if they cannot monitor my actions.[a]	0.735	
Management believes that employees can be trusted.[a]	0.688	
Trust in technology—reliability		.95
The cybersecurity software at my workplace (eg, antivirus and firewall) is reliable.[a]	0.897	
The cybersecurity software at my workplace does not fail me.[a]	0.939	
The cybersecurity software at my workplace provides accurate service.[a]	0.893	
Trust in technology—functionality		.95
The cybersecurity software at my workplace has the functionality I need.[a]	0.946	
The cybersecurity software at my workplace has the features required for my tasks.[a]	0.929	
The cybersecurity software at my workplace has the ability to do what I want it to do.[a]	0.909	
Perceived information security risk		.93
At my workplace, the risk to my computer and data from Internet security breaches is[d]:	0.704	
At my workplace, the likelihood that my computer will be disrupted due to Internet security breaches within the next 12 months is[d]:	0.918	
At my workplace, the chance that my computer will fall a victim to an Internet security breach is[d]:	0.967	
At my workplace, the vulnerability of my computer and data to Internet security risks is[d]:	0.910	
Workload		.82
I feel that the number of requests, problems, or complaints I deal with at work is more than expected.[a]	Dropped	

Construct and items	Loadings	Cronbach alpha
I feel that the amount of work I do interferes with how well it is done.[a]	0.588	
I feel busy or rushed at work. (R)[e]	0.916	
I feel pressured at work. (R)[e]	0.818	

[a]Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree.

[b]Not applicable.

[c]Single-item measurement; strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, strongly disagree.

[d]Extremely high, somewhat high, neither high nor low, somewhat low, or extremely low.

[e](R): Reverse coded item; always, most of the time, about half the time, sometimes, never.

As hospitals' IT departments knew the identity of clickers, they distributed 1 survey link to employees that had clicked on the phishing link and the other survey link to those that had not. This separation helped facilitate the anonymity of the survey analysis, as we did not ask hospitals for any clicking data. Participants were informed that their participation in our survey was voluntary and anonymous.

By combining these 2 steps, we aimed to independently and systematically assess the extent to which attitudes and attributes are related to the actual clicking behavior. Collecting clicking data in the first step has the advantage that the results are not distorted by the survey. To nullify the concern of whether having clicked or not clicked on the phishing email would influence behavior in the survey, the survey was distributed about 6 weeks after the phishing emails were sent out.

Measures

The survey contained questions about the personal attitudes toward the company and its ISP. A pilot test was run with 10 researchers to ensure that all questions were clear. A 5-point Likert scale (1=never, strongly disagree, or extremely low; and 5=always, strongly agree, or extremely high) was used for all items except for intention, our only single-item measurement, where a 7-point Likert scale (1=strongly disagree and 7=strongly agree) was used. See recommendations by Fuchs and Diamantopoulos [47] and Wirtz and Lee [48] for using a larger scale for single-item measurements.

All survey items were based on previously validated items to maximize reliability. The 9 constructs of the survey include attitudes, subjective norm, perceived behavioral control, intention [8,25], collective felt trust [29,40], trust in security technology based on reliability and functionality [22,41], perceived security risk [49], and workload [50].

As control variables in addition to the core concepts, we also asked for the average number of emails received daily, age, position (clinical or nonclinical), and education level.

Data Analysis

The survey was sent to 3169 employees in 3 hospital networks. A total of 488 individuals participated in the study (488/3169, 15.40% response rate). Owing to missing data in variables essential for the proposed research model (eg, *Intention to comply*), 58 participants were excluded from the analysis. To minimize external influences, we also excluded participants from hospital network C, because from this hospital (a small local hospital), only 33 employees participated in the survey. Together, this led to the exclusion of 91 participants and a final sample of 397 (397/3100, 12.80% overall response rate). Of the remaining participants, 172 were from hospital network A, and 225 were from hospital network B.

Table 2 presents the individual response rates, and Table 3 presents the respondent characteristics of the final sample. The respondent characteristics show that the sample is heterogeneous, having a positive impact on the external validity of this study (see Table 3).

As a proxy to test for nonresponse bias, we followed the recommendations by Armstrong and Overton [51] and tested differences in age, gender, position, education, and clicking behavior between early and late respondents. *t* test results show no significant differences between these 2 groups.

To test the strength of the relationship among different constructs and its effect on the actual clicking behavior, we used partial least squared structural equation modeling (SEM) in software SmartPLS (SmartPLS GmbH). PLS was chosen over covariance-based SEM, as it is widely applied in information security research [52] and does not assume a normal distribution, is particularly appropriate for complex models, and its bootstrapping method increases robustness [53].

Before testing the SEM, we assessed the constructs' loadings and Cronbach alphas to evaluate the reliability of the measurement model. After 2 items were dropped from the analysis, all loadings were above the common threshold value of 0.70 [53]. In addition, Cronbach alphas all exceeded .70, also indicating good reliability [53]. Furthermore, the constructs showed adequate convergent validity as the average variance extracted (AVE) was above 0.68 and composite reliability was above 0.70 for all factors.

Discriminant validity was tested by using the Fornell-Lacker Criterion and Heterotrait-Monotrait ratios. The Fornell-Lacker Criterion indicated that the square root of the AVE of each construct was higher than the construct's correlation with any other construct [54]. In addition, Heterotrait-Monotrait ratios were below the threshold of 0.9, also confirming discriminant validity for the measurement models [55]. Finally, all variance inflation factors values were below 5, which suggests multicollinearity is unlikely to be a problem [53]. The relevant reliability and validity fit indices on construct level are reported in Table 4.

Table 2. Response rates.

Hospital network and target group	Employees who received the questionnaire (N)	Responses included in the analysis, n (%)
Hospital network A		
Total	2100	172 (8.20)
Clicker	1600	122 (7.63)
Nonclicker	500	50 (10.0)
Hospital network B		
Total	1000	225 (22.50)
Clicker	500	109 (21.8)
Nonclicker	500	116 (23.2)
Overall sample total	3100	397 (12.80)

Table 3. Respondent characteristics (N=397).

Category	Count, n (%)
Sex	
Male	82 (22.09)
Female	309 (76.28)
Nonbinary	2 (0.47)
Unanswered	4 (1.16)
Age (years)	
18-24	28 (7.05)
25-34	108 (27.20)
35-44	70 (17.63)
45-54	78 (19.65)
55-64	86 (21.66)
65-74	19 (4.79)
≥75	2 (0.50)
Unanswered	6 (1.51)
Position	
Clinical	221 (55.67)
Nonclinical	172 (43.32)
Unanswered	4 (1.01)
Education	
Less than high school	28 (7.30)
High school graduate	47 (11.84)
Some college	111 (27.96)
2-year degree	43 (10.83)
4-year degree	120 (30.23)
Professional degree	41 (10.33)
Unanswered	6 (1.51)
Emails per day	
<10	87 (21.91)
11-20	133 (33.50)
21-30	72 (18.14)
>31	101 (25.44)
Unanswered	4 (1.01)
Response to phishing email	
Clicker	231 (58.19)
Nonclicker	166 (41.81)

Table 4. Reliability and validity of measurement model.

Construct	Cronbach alpha	Average variance extracted	Composite reliability	Heterotrait-Monotrait ratio							
				Attitudes	Subjective norm	Perceived behavioral control	Intention to comply	Collective felt trust	Reliability	Functionality	Perceived risk
Attitudes	.88	0.80	0.92	__[a]	—	—	—	—	—	—	—
Subjective norm	.94	0.89	0.96	0.391	—	—	—	—	—	—	—
Perceived behavioral control	.84	0.76	0.90	0.419	0.381	—	—	—	—	—	—
Intention to comply	N/A[b]	N/A	N/A	0.486	0.337	0.621	—	—	—	—	—
Collective felt trust	.76	0.69	0.87	0.270	0.208	0.270	0.289	—	—	—	—
Reliability	.94	0.89	0.96	0.298	0.251	0.320	0.466	0.324	—	—	—
Functionality	.95	0.91	0.97	0.289	0.231	0.510	0.382	0.351	0.871	—	—
Perceived risk	.93	0.83	0.95	0.117	0.165	0.252	0.270	0.299	0.320	0.196	—
Workload	.81	0.73	0.89	0.122	0.032	0.224	0.146	0.219	0.161	0.188	0.178

[a]Table is symmetric, only the lower triangle is presented.

[b]N/A: not applicable.

Results

Table 5 reports means, SDs, and zero-order correlations of all latent variables. For testing the SEM, bias-corrected bootstrapping based on a bootstrap sample of 5000 was applied [53]. The standardized paths coefficients and significance levels are presented in Figure 2.

Hypothesis 1 predicted that (1) attitudes toward ISP, (2) subjective norms, and (3) perceived behavioral control are positively related to the intention to comply. This prediction is supported: attitudes toward ISP (beta=.27; $P<.001$), subjective norm (beta=.08; $P=.04$), and perceived behavioral control (beta=.44; $P<.001$) showed significant relationships with intention to comply.

Hypothesis 2 predicted that the intention to comply is positively related to compliance behavior. Contrary to the assumption, our results show that intention and clicking behavior are not significantly related in our analysis (beta=−.03; $P=.57$). Thus, hypothesis 2 is not supported.

Hypothesis 3 predicted that collective felt trust is positively related to attitudes toward ISP and subjective norms. Our results support this hypothesis: Trust is significantly related to attitudes toward ISP (beta=.23; $P<.001$) and subjective norm (beta=.18; $P=.001$).

Hypothesis 4 predicted that trust in security technology, consisting of (1) reliability and (2) functionality, is positively related to perceived behavioral control. Our results support hypothesis 4a (beta=.42; $P<.001$) but not 4b (beta=.11; $P=.15$). Thus, trust in security technology is solely based on reliability perceptions.

Hypothesis 5 predicted that a higher perceived risk of cyberattacks is negatively related to the likelihood to click on phishing links. This hypothesis cannot be supported as our results indicate no significant relationship between perceived risk and the behavior to click on phishing links (beta=.10; $P=.05$).

Finally, hypothesis 6 predicted that high workload is positively related to the likelihood of clicking on phishing links. Our results show that this relationship is indeed significant, supporting this hypothesis (beta=.16; $P=.001$).

We included several control variables to test whether clicking on a phishing link was different for age groups, education levels, positions (clinical or nonclinical), or the number of emails received per day. None of these variables had a significant effect on the behavior of clicking on the link in the phishing email.

Table 5. Zero-order correlations and descriptive statistics.

Construct	Value, mean (SD)	Zero-order correlations							
		Attitudes	Subjective norm	Perceived behavioral control	Intention to comply	Collective felt trust	Reliability	Functionality	Perceived risk
Attitudes	4.79 (0.42)	—[a]	—	—	—	—	—	—	—
Subjective norm	4.42 (0.72)	.38[b]	—	—	—	—	—	—	—
Perceived behavioral control	4.46 (0.38)	.34[b]	.34[b]	—	—	—	—	—	—
Intention to comply	6.69 (0.572)	.47[b]	.34[b]	.58[b]	—	—	—	—	—
Collective felt trust	4.81 (0.88)	.26[b]	.21[b]	.23[b]	.29[b]	—	—	—	—
Reliability	4.09 (0.75)	.28[b]	.25[b]	.55[b]	.46[b]	.32[b]	—	—	—
Functionality	4.05 (0.92)	.27[b]	.22[b]	.49[b]	.38[b]	.34[b]	.87[b]	—	—
Perceived risk	2.46 (0.84)	−.10	−.16[c]	−.24[b]	−.26[b]	−.29[b]	−.32[b]	−.20[b]	
Workload	2.76 (0.72)	−.11	.01	−.18[c]	−.12[d]	−.19[c]	−.14[d]	−.17[c]	.16[c]

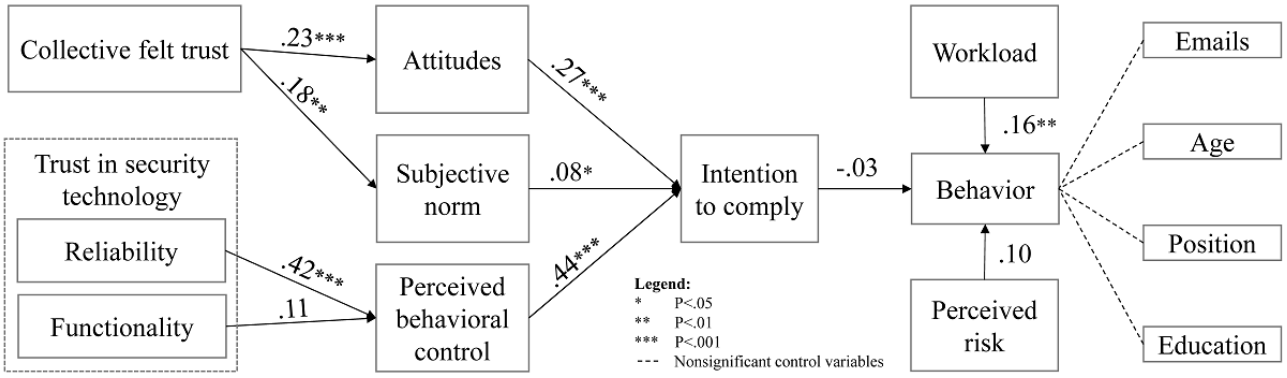
[a]Table is symmetric, only the lower triangle is presented.

[b]*P*<.001, 2-tailed.

[c]*P*<.01, 2-tailed.

[d]*P*<.05, 2-tailed.

Figure 2. Results of structural equation model.



Evaluating the overall model, we find that the data fit well for intention to comply with an R^2 of 0.397. The Stone-Geisser test has shown with a Q^2 value of 0.377 that the model has a large predictive relevance for intention to comply. In contrast, the model explains little variance of the clicking behavior (R^2=0.044) and has a small predictive relevance for this construct (Q^2=0.040).

Employees from 2 hospital networks were included in this analysis, with 172 participants being employed by hospital network A and 225 by hospital network B. Performing a subgroup analysis, we determined whether the hospital context affected the SEM results. Results of the subgroup analysis show no group differences regarding the (non) significant paths (see Table 6). Hypothesis 1b is only partially supported, as the overall effect is significant here, although no significant effect can be identified in hospitals A and B.

A multigroup analysis (MGA) was performed to test whether any significant differences in these path coefficients exist. The test for measurement invariance shows no significant differences in the measurement models between hospital networks A and B, which indicates that potential differences are not based on measurement error and that MGA provides reliable results at the construct level. In the overall sample, no control variable has a significant effect on the clicking behavior. However, the MGA reveals that the position (clinical vs nonclinical) significantly affects the clicking behavior in hospital B and that this effect differs significantly from the effect in hospital A (|hospital A-B|=0.216; *P*>.99). On the other hand, in hospital A, education has a significantly negative effect on the clicking behavior. This effect is not observed in hospital B and is significantly different (|hospital A-B|=0.184; *P*=.99).

Table 6. Results of structural equation model and its multi-group analysis.

Hypotheses	Overall sample		Hospital A		Hospital B		Multigroup analysis		Assessment of hypotheses
	Beta[a]	*P* value	Beta	*P* value	Beta	*P* value	|Difference|	*P* value	
H1a	.268	<.001	.367	<.001	.172	.047	.195	.11	Supported
H1b	.083	.04	.105	.09	.053	.32	.052	.52	Partly supported
H1c	.444	<.001	.403	<.001	.490	<.001	.087	.40	Supported
H2	−.037	.45	−.041	.58	−.021	.76	.020	.85	Rejected
H3a	.229	<.001	.238	<.001	.234	<.001	.004	.97	Supported
H3b	.178	<.001	.179	.02	.178	.02	.001	.99	Supported
H4a	.421	<.001	.424	<.001	.435	<.001	.011	.95	Supported
H4b	.112	.15	.144	.20	.087	.41	.057	.71	Rejected
H5	.099	.05	.051	.58	.091	.22	.040	.73	Rejected
H6	.157	<.001	.242	<.001	.137	.04	.105	.26	Supported
Controls									
Emails	.063	.26	−.071	.39	.112	.10	.183	.93	—[b]
Age	.013	.81	−.070	.36	.027	.72	.097	.82	—[b]
Position	.076	.14	−.083	.10	.133	<.001	.216	>.99	—[b]
Education	.018	.74	−.108	.01	.076	.25	.184	.99	—[b]

[a]Beta=effect size.

[b]Not applicable.

Discussion

Principal Findings

This study investigates the relationship between employees' compliance intention and their actual compliance with ISPs (ie, not clicking on the phishing link). As hypothesized in H1, we found that attitudes, subjective norm, and perceived behavioral control were positively related to the intention to comply with organizational ISPs. However, contrary to what was suggested in H2, there was no significant relationship between the intention to comply and the compliance behavior itself. In contrast to this finding, previous studies have provided evidence for a positive relationship [56,57]. However, because of the difficulty of observing actual behavior, these studies have relied on self-reported data to assess the relationship between intention and the actual compliance. This process leaves room for method biases because individuals could give desirable answers, or previous answers could influence later answers [58,59]. In a recent review of employees' security behavior, the authors challenge the assumption that intention predicts behavior in the information security context [15]. In line with this, our results indicate that in the context of phishing emails, intention and compliance might not be as strongly linked as previously assumed. Thus, the role of context in compliance investigations should be carefully considered as it might prove to be highly relevant.

We also found that collective felt trust was significantly related to employees' attitudes and subjective norms, supporting H3. Higher collective felt trust is associated with more positive attitudes and subjective norms, which in turn positively influence the compliance intention. The results indicate that management can have an influence on how employees perceive security policies. Moreover, the rich literature of trust and control points toward another benefit: Trust in the management reduces the risk that employees perceive security policies as a sign of management distrust in them and their abilities [60]—employees might understand that ISP are not designed to reduce their freedom but to enhance their protection. In addition, with high levels of trust, employees are likely to internalize the organization's goals and thus are more willing to protect the company by accepting the policies [61,62].

Considering the relationship between trust in technology and perceived behavioral control, we found that only trust in technology based on reliability has a significant influence on an employee's perceived behavioral control. Trust based on functionality was not significantly related. This finding supports H4a but not H4b. With high trust in information security technologies in use, employees may think that they are more capable of controlling their own behavior. Trust in technology has been shown to increase the adoption of new technologies and has frequently been used in IS literature [63,64]. Such reliance and trust in technology is also seen in health care settings as medical professions constantly introduce or utilize programs that allow most convenient—as well as easy—access to patient records [65]. This is done to ensure that physicians can offer the best care for their patients. Unbeknown to them, however, this also allows for easier access to sensitive personal information for attackers. Furthermore, research on information system security has not yet been adopted into this concept. Thus,

these findings not only contribute to security compliance but also enhance the understanding of application areas for trust in technology.

Moreover, we found no significant effect of perceived risk on compliance behavior, contrary to what was assumed in H5. A reason for this might be that the risk is too abstract for employees or that the perceived benefits in a certain situation outweigh the perceived risk [49]. More specifically, in the health care setting, the risk of clicking on a foreign email or revealing sensitive information would most likely outweigh the risk of patient safety, treatment quality, private information, and data theft—which would be the most plausible explanation for this result [66].

As assumed in H6, we found a significant effect between workload and compliance behavior. As none of the cognitive variables showed a significant relationship with the behavior, the workload is the only variable related to the compliance behavior. This finding is interesting because it offers an insight into the situations in which phishing emails are opened. Any form of noncompliance behavior (in this case, the necessity to cope with a high workload) might lead to less eagerness to follow security policies [12]. Furthermore, high workload might cause unintended noncompliance behavior—high volumes of work could make one click on a phishing link because an overworked employee could have been too occupied to notice the imposed threats [15]. This is especially concerning given that cyberattacks today are extremely hard to detect because they have become extremely intricate; they are targeted attacks that have been carefully planned according to each organization's needs [67]. Tactics such as social engineering—the act of psychologically manipulating people into revealing personal information or allowing access to a secured server—have been increasingly successful in phishing [68]. More specifically, spear-phishing, a specified type of attack, uses context-specific, sophisticated emails that are tailored to meet individual and company-specific needs. It is difficult to detect—and requires much attention to detect—because the reader must consider the *plausibility* of the written text—rather than visual or auditory deception [69]. In the case of an overworked, occupied employee, such sophisticated attacks are more likely to be successful.

Practical Implications

Our findings offer a number of practical implications. Practitioners need to consider organizational factors when designing security policies and training programs. Our results show that engaging in trust-building activities can subsequently enhance employee's compliance. Our findings also highlight the relevance of top management participation and imply that managers need to show that they acknowledge the problems associated with IS security and are able to provide a foundation of security policy and behavior upon which employees can build [8,70,71].

Furthermore, the positive relationship between trust in technology and perceived behavioral control indicates that the feeling of reliance on technology is associated with a higher intention to comply. Besides ensuring good quality of security technologies, managers need to communicate and inform employees about security technologies. If employees cannot learn about the technology in place, they cannot know how much to rely on it. Trust in technology can be developed through training and by enhancing understanding of the technology—see Puhakinen and Siponen [70] and Safa et al [72] for more discussion.

As our results show that in the context of phishing emails, the compliance intention was not related to the actual compliance behavior, hospitals must remain vigilant with vulnerabilities that cannot be easily managed.

Finally, our results present a relationship between workload and noncompliance behaviors. This finding suggests that hospitals should better manage workload to increase information security—for instance, extensive emailing may unnecessarily add to workload. Our observations working with organizations show that in addition to communication with colleagues through emails, individuals receive multiple emails on a daily basis including announcements and other general notes, which add to individuals' email loads, putting them in more risks of clicking on phishing emails.

Limitations

Although this study provides several insights, it is also subject to some limitations. First, the low response rate and the gender imbalance in our sample might indicate a selection bias. Selection bias is often associated with low generalizability of the results, as it is assumed that only a certain group of people responded to a study. Previous research has reported that response rates are generally low in this field [73,74]. To investigate potential bias that arise from this issue, we checked for nonresponse bias via factor analysis using the principal component as well as marker variable test [59,75,76]. All tests showed nonsignificant results, and although the presence of such bias can never be completely excluded, the results suggest that bias is not an issue in our analysis. We also included gender as a control variable in our model; however, the results showed no significant influence. As gender does not explain additional variance, we excluded this control variable from the analysis.

Although our results provide evidence for an insignificant relationship between intention and behavior, future research should investigate this relationship in a different context to validate these findings. We used a specific case of phishing emails to investigate employees' compliance in hospitals. As previous studies have shown, the effects between TPB constructs and influencing variables might depend on the underlying scenario. Moody et al [17] found support for TPB in scenarios concerning USB use, workstation logout, and password sharing. The intention-behavior gap might be more relevant in certain situations than in others. For instance, employees might not intend to open a suspicious email but then end up doing it because of spontaneous curiosity or inattentiveness. Moreover, we focused on the hospital industry to keep the sample as consistent as possible. This restriction might limit the generalizability of our results. Organizational factors and governance structure should be considered when transmitting these results to other industry settings.

Finally, we used a generic measure to assess the intention to comply with ISP, which means we asked for general compliance instead of focusing on phishing emails. We did so because (1) we were interested in the general assessment of their own intention to comply and (2) we did not want to influence the response by drawing attention toward phishing specifically. We think that a generic measure is justified in this situation because employees in the investigated hospitals are expected to know about phishing email regulations. In both hospital networks we investigated, information security staff had already raised awareness of this issue among the employees—both hospital networks had antiphishing email training. Therefore, we did not want to draw additional attention to this matter but pose a broader question about general compliance.

Conclusions

Employees' compliance with policies is a key concern in information security research, especially because even an accidental security breach from a phishing attack could severely impair sensitive patient information and safety in a health care setting. This study focused on factors related to employees' compliance intention and their actual compliance (ie, not clicking on phishing links) in hospitals. Using the lens of the TPB, we investigated the role of collective felt trust and trust in technology as an influence on attitudes, subjective norms, and perceived behavioral control. We found positive effects between collective felt trust and attitudes toward compliance and subjective norms. Trust in technology was strongly related to perceived behavioral control. In the health care setting, this trust is even more evident and exploited, as previous research revealed that more than half of the information breaches came from within the health care organizations. In addition, employees showed strong preference for trust in electronic records—over paper records—because online systems were easily accessible remotely or offsite and readily available in cases of emergency [77].

Surprisingly, we did not find an association between the intention to comply and the actual compliance behavior. With major improvements in cyberattack technologies (ie, tailoring information to be specific to the needs of the target audience or organization), it becomes nearly impossible for both employees and servers to filter phishing emails, and so employees with high intention to comply may still fall in the trap of the hackers. However, we found that a higher workload was positively related to noncompliance behavior. This finding suggests that, in the context of phishing emails, context effects are highly relevant.

A major strength of this study is that we separated data collection for the dependent (clicking behavior) and the independent (personal and organizational) variables. This is one of the few studies in information security literature that observed the compliance behavior rather than using self-reported data. This approach enabled us to obtain more reliable outcomes, given that self-reports may differ based on perception and mood, among others. We hope that our findings motivate the information security community to improve current training programs and design effective interventions to increase information security compliance.

Acknowledgments

The authors would like to express their gratitude to chief security officers at the anonymous hospital networks who greatly contributed to this study and facilitated their data collection process. They also wish to thank all those who gave their time to participate in their survey. Financial support for this study was provided by Cybersecurity at MIT Sloan (CAMS), also known as MIT (IC)^3—the Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity. MSJ was supported in the preparation of this research by CAMS. MB, DW, and GS were supported in part by a grant from the German Research Foundation (grant no. 1712/2).

Conflicts of Interest

None declared.

References

1. Coventry L, Branley D. Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas 2018 Jul;113:48-52. [doi: 10.1016/j.maturitas.2018.04.008] [Medline: 29903648]
2. Jalali MS, Kaiser JP. Cybersecurity in hospitals: a systematic, organizational perspective. J Med Internet Res 2018 May 28;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
3. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. Health care and cybersecurity: bibliometric analysis of the literature. J Med Internet Res 2019 Feb 15;21(2):e12644 [FREE Full text] [doi: 10.2196/12644] [Medline: 30767908]
4. Jalali MS, Russell B, Razak S, Gordon W. EARS to cyber incidents in health care. J Am Med Inform Assoc 2019 Jan 1;26(1):81-90. [doi: 10.1093/jamia/ocy148] [Medline: 30517701]
5. Gordon WJ, Wright A, Aiyagari R, Corbo L, Glynn RJ, Kadakia J, et al. Assessment of employee susceptibility to phishing attacks at US health care institutions. JAMA Netw Open 2019 Mar 1;2(3):e190393 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0393] [Medline: 30848810]
6. Kruse CS, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]

7. Boss SR, Kirsch LJ, Angermeier I, Shingler RA, Boss RW. If someone is watching, I'll do what I'm asked: mandatoriness, control, and information security. Eur J Inf Syst 2017;18(2):151-164. [doi: 10.1057/ejis.2009.8]
8. Hu Q, Dinev T, Hart P, Cooke D. Managing employee compliance with information security policies: the critical role of top management and organizational culture*. Decis Sci 2012;43(4):615-660. [doi: 10.1111/j.1540-5915.2012.00361.x]
9. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. Manag Inf Syst Q 2010 Sep;34(3):523-548. [doi: 10.2307/25750690]
10. Herath T, Rao HR. Encouraging information security behaviors in organizations: role of penalties, pressures and perceived effectiveness. Decis Support Syst 2009;47(2):154-165. [doi: 10.1016/j.dss.2009.02.005]
11. Johnston AC, Warkentin M. Fear appeals and information security behaviors: an empirical study. Manag Inf Syst Q 2010;34(3):549-566. [doi: 10.2307/25750691]
12. Siponen M, Vance A. Neutralization: new insights into the problem of employee information systems security policy violations. Manag Inf Syst Q 2010 Sep;34(3):487-502. [doi: 10.2307/25750688]
13. Gordon WJ, Wright A, Glynn RJ, Kadakia J, Mazzone C, Leinbach E, et al. Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. J Am Med Inform Assoc 2019 Jun 1;26(6):547-552. [doi: 10.1093/jamia/ocz005] [Medline: 30861069]
14. Uwizeyemungu S, Poba-Nzaou P, Cantinotti M. European hospitals' transition toward fully electronic-based systems: do information technology security and privacy practices follow? JMIR Med Inform 2019 Mar 25;7(1):e11211 [FREE Full text] [doi: 10.2196/11211] [Medline: 30907732]
15. Lebek B, Uffen J, Neumann M, Hohler B, Breitner MH. Information security awareness and behavior: a theory-based literature review. Manag Res Rev 2014;37(12):1049-1092. [doi: 10.1108/mrr-04-2013-0085]
16. Siponen M, Mahmood MA, Pahnila S. Employees' adherence to information security policies: an exploratory field study. Inf Manag 2014;51(2):217-224. [doi: 10.1016/j.im.2013.08.006]
17. Moody GD, Siponen M, Pahnila S. Toward a unified model of information security policy compliance. Manag Inf Syst Q 2018;42(1):285-311. [doi: 10.25300/misq/2018/13853]
18. Sommestad T, Hallberg J, Lundholm K, Bengtsson J. Variables influencing information security policy compliance. Inf Manag Comp Security 2014;22(1):42-75. [doi: 10.1108/imcs-08-2012-0045]
19. Fulmer CA, Gelfand MJ. At What Level (and in Whom) We Trust. J Manag 2012;38(4):1167-1230. [doi: 10.1177/0149206312439327]
20. Talaulicar T, Grundei J, Werder AV. Strategic decision making in start-ups: the effect of top management team organization and processes on speed and comprehensiveness. J Bus Ventur 2005;20(4):519-541. [doi: 10.1016/j.jbusvent.2004.02.001]
21. Jarvenpaa SL, Shaw TR, Staples DS. Toward contextualized theories of trust: the role of trust in global virtual teams. Inf Syst Res 2004;15(3):250-267. [doi: 10.1287/isre.1040.0028]
22. Lankton N, McKnight DH, Tripp J. Technology, humanness, and trust: rethinking trust in technology. J Assoc Inf Syst 2015;16(10):880-918. [doi: 10.17705/1jais.00411]
23. McEvily B, Tortoriello M. Measuring trust in organisational research: review and recommendations. J Trust Res 2011;1(1):23-63. [doi: 10.1080/21515581.2011.552424]
24. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. Acad Manage Rev 1998;23(3):393-404. [doi: 10.5465/amr.1998.926617]
25. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Pearson; 1980.
26. Fishbein M, Ajzen I. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Boston, MA: Addison-Wesley; 1975.
27. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
28. Ifinedo P. Understanding information systems security policy compliance: an integration of the theory of planned behavior and the protection motivation theory. Comput Secur 2012;31(1):83-95. [doi: 10.1016/j.cose.2011.10.007]
29. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995;20(3):709-734. [doi: 10.2307/258792]
30. Pavlou PA, Fygenson M. Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. Manag Inf Syst Q 2006;30(1):115-143. [doi: 10.2307/25148720]
31. Chen ZX, Aryee S, Lee C. Test of a mediation model of perceived organizational support. J Vocat Behav 2005;66(3):457-470. [doi: 10.1016/j.jvb.2004.01.001]
32. Whitener EM. Do 'high commitment' human resource practices affect employee commitment? A cross-level analysis using hierarchical linear modeling. J Manag 2001;27(5):515-535. [doi: 10.1177/014920630102700502]
33. Krosgaard MA, Brodt SE, Whitener EM. Trust in the face of conflict: the role of managerial trustworthy behavior and organizational context. J Appl Psychol 2002;87(2):312-319. [doi: 10.1037//0021-9010.87.2.312]
34. Mayer RC, Gavin MB. Trust in management and performance: who minds the shop while the employees watch the boss? Acad Manag J 2005;48(5):874-888. [doi: 10.5465/amj.2005.18803928]
35. Dirks KT, Ferrin DL. Trust in leadership: meta-analytic findings and implications for research and practice. J Appl Psychol 2002;87(4):611-628. [doi: 10.1037//0021-9010.87.4.611]

36. Colquitt JA, Scott BA, LePine JA. Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. J Appl Psychol 2007 Jul;92(4):909-927. [doi: 10.1037/0021-9010.92.4.909] [Medline: 17638454]
37. Davies MA, Lassar W, Manolis C, Prince M, Winsor RD. A model of trust and compliance in franchise relationships. J Bus Ventur 2011;26(3):321-340. [doi: 10.1016/j.jbusvent.2009.09.005]
38. Kim WC, Mauborgne RA. Procedural Justice, Attitudes, and Subsidiary Top Management Compliance with Multinationals' Corporate Strategic Decisions. Acad Manag J 1993;36(3):502-526. [doi: 10.2307/256590]
39. Rafaeli A, Sagy Y, Derfler-Rozin R. Logos and initial compliance: a strong case of mindless trust. Organ Sci 2008;19(6):845-859. [doi: 10.1287/orsc.1070.0344]
40. Salamon SD, Robinson SL. Trust that binds: the impact of collective felt trust on organizational performance. J Appl Psychol 2008 May;93(3):593-601. [doi: 10.1037/0021-9010.93.3.593] [Medline: 18457488]
41. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manag Inf Syst 2011;2(2):1-25. [doi: 10.1145/1985347.1985353]
42. Srivastava SC, Chandra S, Shirish A. Technostress creators and job outcomes: theorising the moderating influence of personality traits. Inf Syst J 2015;25(4):355-401. [doi: 10.1111/isj.12067]
43. Hernández-Ortega B. The role of post-use trust in the acceptance of a technology: drivers and consequences. Technovation 2011;31(10-11):523-538. [doi: 10.1016/j.technovation.2011.07.001]
44. Vance A, Elie-Dit-Cosaque C, Straub DW. Examining trust in information technology artifacts: the effects of system quality and culture. J Manag Inf Syst 2014;24(4):73-100. [doi: 10.2753/mis0742-1222240403]
45. Ho SM, Ocasio-Velázquez M, Booth C. Trust or consequences? Causal effects of perceived risk and subjective norms on cloud technology adoption. Comput Secur 2017;70:581-595. [doi: 10.1016/j.cose.2017.08.004]
46. Han W, Ada S, Sharman R, Rao HR. Campus emergency notification systems: an examination of factors affecting compliance with alerts. Manag Inf Syst Q 2015;39(4):909-929. [doi: 10.25300/MISQ/2015/39.4.8]
47. Fuchs C, Diamantopoulos A. Using single-item measures for construct measurement in management research: conceptual issues and application guidelines. Die Betriebswirtschaft 2009;69(2):195-210 [FREE Full text]
48. Wirtz J, Lee MC. An examination of the quality and context-specific applicability of commonly used customer satisfaction measures. J Serv Res 2003;5(4):345-355. [doi: 10.1177/1094670503005004006]
49. Li H, Zhang J, Sarathy R. Understanding compliance with internet use policy from the perspective of rational choice theory. Decis Support Syst 2010;48(4):635-645. [doi: 10.1016/j.dss.2009.12.005]
50. Rutner PS, Hardgrave BC, McKnight DH. Emotional dissonance and the information technology professional. Manag Inf Syst Q 2008;32(3):635-652. [doi: 10.2307/25148859]
51. Armstrong JS, Overton TS. Estimating nonresponse bias in mail surveys. J Mark Res 1977;14(3):396-402. [doi: 10.2307/3150783]
52. Ringle CM, Sarstedt M, Straub DW. Editor's Comments: A Critical Look at the Use of PLS-SEM in 'MIS Quarterly'. Manag Inf Syst Q 2012;36(1):iii-xiv. [doi: 10.2307/41410402]
53. Hair JF, Hult GT, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). London: Sage publications; 2016.
54. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981;18(1):39-50. [doi: 10.1177/002224378101800104]
55. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2015;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
56. Myyry L, Siponen M, Pahnila S, Vartiainen T, Vance A. What levels of moral reasoning and values explain adherence to information security rules? An empirical study. Eur J Inf Syst 2017;18(2):126-139. [doi: 10.1057/ejis.2009.10]
57. Pahnila S, Siponen M, Mahmood A. Employees' Behavior towards IS Security Policy Compliance. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences.: IEEE Computer Society; 2007 Presented at: HICSS'07; January 3-6, 2007; Hawaii. [doi: 10.1109/hicss.2007.206]
58. Jalali MS. How individuals weigh their previous estimates to make a new estimate in the presence or absence of social influence. In: Kennedy WG, Agarwal N, Yang SJ, editors. Social Computing, Behavioral-Cultural Modeling and Prediction. Cham: Springer International Publishing; 2014:67-74.
59. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
60. Weibel A, den Hartog DN, Gillespie N, Searle R, Six F, Skinner D. How do controls impact employee trust in the employer? Hum Resour Manag 2015;55(3):437-462. [doi: 10.1002/hrm.21733]
61. Deery SJ, Iverson RD, Walsh JT. Toward a better understanding of psychological contract breach: a study of customer service employees. J Appl Psychol 2006 Jan;91(1):166-175. [doi: 10.1037/0021-9010.91.1.166] [Medline: 16435946]
62. Maguire S, Phillips N. 'Citibankers' at Citigroup: a study of the loss of institutional trust after a merger. J Manag Stud 2008;45(2):372-401. [doi: 10.1111/j.1467-6486.2007.00760.x]

63. Lankton NK, McKnight DH, Wright RT, Thatcher JB. Research Note—using Expectation Disconfirmation Theory and Polynomial Modeling to understand trust in technology. Inf Syst Res 2016;27(1):197-213. [doi: 10.1287/isre.2015.0611]
64. Lowry PB, Zhang D, Zhou L, Fu X. Effects of culture, social presence, and group composition on trust in technology-supported decision-making groups. Inf Syst J 2010;20(3):297-315. [doi: 10.1111/j.1365-2575.2009.00334.x]
65. Shenoy A, Appel JM. Safeguarding confidentiality in electronic health records. Camb Q Healthc Ethics 2017 Apr;26(2):337-341. [doi: 10.1017/S0963180116000931] [Medline: 28361730]
66. Pullin DW. Cybersecurity: positive changes through processes and team culture. Front Health Serv Manage 2018;35(1):3-12. [doi: 10.1097/HAP.0000000000000038] [Medline: 30124509]
67. Rajivan P, Gonzalez C. Creative persuasion: a study on adversarial behaviors and strategies in phishing attacks. Front Psychol 2018;9:135 [FREE Full text] [doi: 10.3389/fpsyg.2018.00135] [Medline: 29515478]
68. Anderson RJ. Security Engineering: A Guide to Building Dependable Distributed Systems. Hoboken, NJ: Wiley; 2013.
69. Schütz SW, Lowry PB, Thatcher JB. Defending against Spear Phishing: Motivating Users through Fear appeal Manipulations. In: Proceedings of the 2016 Pacific Asia Conference on Information Systems. 2016 Presented at: PACIS'16; June 27-July 1, 2016; Chiayi City, Taiwan.
70. Puhakainen P, Siponen M. Improving employees' compliance through information systems security training: an action research study. Manag Inf Syst Q 2010;34(4):757-778. [doi: 10.2307/25750704]
71. Kankanhalli A, Teo H, Tan BC, Wei K. An integrative study of information systems security effectiveness. Int J Inf Manag 2003;23(2):139-154. [doi: 10.1016/S0268-4012(02)00105-6]
72. Safa NS, von Solms R, Furnell S. Information security policy compliance model in organizations. Comput Secur 2016;56:70-82. [doi: 10.1016/j.cose.2015.10.006]
73. Guo KH, Yuan Y, Archer NP, Connelly CE. Understanding nonmalicious security violations in the workplace: a composite behavior model. J Manag Inf Syst 2014;28(2):203-236. [doi: 10.2753/mis0742-1222280208]
74. Kotulic AG, Clark JG. Why there aren't more information security research studies. Inf Manag 2004;41(5):597-607. [doi: 10.1016/j.im.2003.08.001]
75. Liang H, Saraf N, Hu Q, Xue Y. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. Manag Inf Syst Q 2007;31(1):59-87. [doi: 10.2307/25148781]
76. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
77. Filkins BL, Kim JY, Roberts B, Armstrong W, Miller MA, Hultner ML, et al. Privacy and security in the era of digital health: what should translational researchers know and do about it? Am J Transl Res 2016;8(3):1560-1580 [FREE Full text] [Medline: 27186282]

Abbreviations

AVE: average variance extracted
IS: information systems
ISP: information security policies
IT: information and technology
SEM: structural equation modeling
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by M Goebel, L Sbaffi; comments to author 19.11.19; revised version received 11.12.19; accepted 16.12.19; published 23.01.20.

Please cite as:
Jalali MS, Bruckes M, Westmattelmann D, Schewe G
Why Employees (Still) Click on Phishing Links: Investigation in Hospitals
J Med Internet Res 2020;22(1):e16775
URL: http://www.jmir.org/2020/1/e16775/
doi:10.2196/16775
PMID:32012071

Original Paper

The Detection of Opioid Misuse and Heroin Use From Paramedic Response Documentation: Machine Learning for Improved Surveillance

José Tomás Prieto[1], PhD; Kenneth Scott[2], PhD; Dean McEwen[2], MBA; Laura J Podewils[2], PhD; Alia Al-Tayyib[2,3], PhD; James Robinson[4], MA; David Edwards[4], EMT-P; Seth Foldy[2,3,5], MPH, MD; Judith C Shlay[2,5], MSPH, MD; Arthur J Davidson[2,5,6], MSPH, MD

[1]Division of Scientific Education and Professional Development, Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Denver Public Health, Denver Health and Hospital Authority, Denver, CO, United States

[3]Department of Epidemiology, Colorado School of Public Health, Aurora, CO, United States

[4]Denver Health Paramedics, Denver Health and Hospital Authority, Denver, CO, United States

[5]Department of Family Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[6]Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO, United States

Corresponding Author:
José Tomás Prieto, PhD
Division of Scientific Education and Professional Development
Centers for Disease Control and Prevention
1600 Clifton Rd
Atlanta, GA, 30333
United States
Phone: 1 3036024487
Email: josetomasprieto@gmail.com

Abstract

Background: Timely, precise, and localized surveillance of nonfatal events is needed to improve response and prevention of opioid-related problems in an evolving opioid crisis in the United States. Records of naloxone administration found in prehospital emergency medical services (EMS) data have helped estimate opioid overdose incidence, including nonhospital, field-treated cases. However, as naloxone is often used by EMS personnel in unconsciousness of unknown cause, attributing naloxone administration to opioid misuse and heroin use (OM) may misclassify events. Better methods are needed to identify OM.

Objective: This study aimed to develop and test a natural language processing method that would improve identification of potential OM from paramedic documentation.

Methods: First, we searched Denver Health paramedic trip reports from August 2017 to April 2018 for keywords naloxone, heroin, and both combined, and we reviewed narratives of identified reports to determine whether they constituted true cases of OM. Then, we used this human classification as reference standard and trained 4 machine learning models (random forest, k-nearest neighbors, support vector machines, and L1-regularized logistic regression). We selected the algorithm that produced the highest area under the receiver operating curve (AUC) for model assessment. Finally, we compared positive predictive value (PPV) of the highest performing machine learning algorithm with PPV of searches of keywords naloxone, heroin, and combination of both in the binary classification of OM in unseen September 2018 data.

Results: In total, 54,359 trip reports were filed from August 2017 to April 2018. Approximately 1.09% (594/54,359) indicated naloxone administration. Among trip reports with reviewer agreement regarding OM in the narrative, 57.6% (292/516) were considered to include information revealing OM. Approximately 1.63% (884/54,359) of all trip reports mentioned heroin in the narrative. Among trip reports with reviewer agreement, 95.5% (784/821) were considered to include information revealing OM. Combined results accounted for 2.39% (1298/54,359) of trip reports. Among trip reports with reviewer agreement, 77.79% (907/1166) were considered to include information consistent with OM. The reference standard used to train and test machine learning models included details of 1166 trip reports. L1-regularized logistic regression was the highest performing algorithm (AUC=0.94; 95% CI 0.91-0.97) in identifying OM. Tested on 5983 unseen reports from September 2018, the keyword naloxone inaccurately identified and underestimated probable OM trip report cases (63 cases; PPV=0.68). The keyword heroin yielded

more cases with improved performance (129 cases; PPV=0.99). Combined keyword and L1-regularized logistic regression classifier further improved performance (146 cases; PPV=0.99).

Conclusions: A machine learning application enhanced the effectiveness of finding OM among documented paramedic field responses. This approach to refining OM surveillance may lead to improved first-responder and public health responses toward prevention of overdoses and other opioid-related problems in US communities.

(J Med Internet Res 2020;22(1):e15645) doi:10.2196/15645

KEYWORDS

naloxone; emergency medical services; natural language processing; heroin; substance-related disorders; opioid crisis; artificial intelligence

Introduction

Background

The more than 47,000 opioid-involved overdose deaths in 2018 in the United States [1,2] insufficiently reflect the nonfatal burden associated with prescription opioid misuse and heroin use (OM) by an estimated 10.3 million people [3]. Timely, precise, and localized surveillance of nonfatal events is needed to define medical treatment trends related to OM and improve response and prevention of overdoses and other opioid-related problems.

Timely information sources about nonfatal opioid-related events include hospitals, emergency departments (EDs) [4], and prehospital emergency medical services (EMS). Paramedics routinely encounter patients with symptoms consistent with drug overdose and administer naloxone (an effective opioid antagonist) to reverse symptoms [5]. EMS data have helped estimate opioid overdose incidence, including nonhospital, field-treated cases [6-8]. Frequency of naloxone administration has positively correlated with opioid and heroin overdose-related ED visits [9] and fatal opioid overdose rates [10], suggesting that naloxone administration might be a relevant proxy to monitor need for interventions.

Opioid misuse and *heroin use* [11] refer to illicit use and nonmedical prescription opioid use for extended periods or for experience and feelings derived from the medication [12]. Naloxone, administered by paramedics to reverse opioid-induced respiratory depression [13,14], might serve as a potential OM sentinel, particularly when OM has resulted in an opioid overdose [5,9,10]. However, as naloxone is often used by EMS personnel in unconsciousness of unknown cause, attributing naloxone administration to opioid overdose and OM may misclassify events as opioid-related. A study of EMS-administered naloxone reported poor sensitivity and low positive predictive value (PPV) for opioid overdose [15].

Objective

Better methods are needed to accurately identify opioid-related problems and trends of OM. To fill this gap, we sought to develop and test a natural language processing (NLP) method that would improve classification of OM among paramedic trip reports with documentation of naloxone administration or evidence of heroin use.

Methods

Setting

Denver Health's (DH) [16] Paramedic Division is the main provider of EMS for the city and county of Denver. Their record system adheres to the National Emergency Medical Services Information System data standard version 3.4.0 [17]. We processed the following variables for each trip report: free-text narratives, primary impressions, alcohol or drug use note, and list of medications administered by paramedics. Table 1 summarizes the 3 study phases.

The Quality Improvement Committee of DH, which is endorsed by the Colorado Multiple Institutional Review Board at the University of Colorado, Denver, determined that this work did not constitute human subjects research.

Table 1. Summary of study phases to classify emergency medical services trip reports for potential opioid misuse, Denver, Colorado, 2017.

Phase	Purpose	Description of methods	Time frame
1	Assess performance of keyword search approaches	Searched trip reports for keywords (ie, "naloxone," "heroin," and both combined) and reviewed charts of identified reports to assess positive predictive value	August 2017 to April 2018
2	Train and test supervised machine learning classification	Guided machine learning models using previous phase's chart review classification results and selected the highest performing algorithm in binary classification of opioid misuse and heroin use	August 2017 to April 2018
3	Validate performance measures across approaches	Compared the highest performing machine learning algorithm with the performance of searches of keywords "naloxone," "heroin," and combination of both	September 2018

Phase 1: Assess Text String Search Approaches

Naloxone administrations have been previously used to flag potential OM resulting in opioid overdoses [5,9,10], and heroin use implies OM. To reduce the DH EMS dataset to a prescreened subset of all paramedic reports, we searched for presence of keywords *naloxone* (or *narcan*) among administered medications or *heroin* (or misspelled variations *herion* and

heroine) in trip report narratives between August 1, 2017, and April 30, 2018. No opioid brand names (eg, Oxycontin or Tramadol) were used to identify opioid-related events. Trip reports that included the keywords were reviewed by 2 independent reviewers, both DH paramedics, to answer the question: "Is there narrative evidence (yes, no or unsure) of illicit opioid use or prescription OM (ie, use beyond clinical needs, for extended periods, or for experience and feelings derived from the medication)?" If unsure or when adverse events from opioids did not imply misuse, reviewers were to classify that report as negative. We hypothesized lower false-positive rates for the *heroin* vs *naloxone* methods because heroin use implies OM. To visualize trends, weekly potential OM paramedic trip report counts for each search approach were calculated. Pearson correlation coefficients (*r*) assessed correlation between weekly OM paramedic trip report counts by search approach and reviewer assessments.

Phase 2: Train and Test Supervised Machine Learning Classification

Trip reports with *naloxone* among administered medications or *heroin* in narratives, plus reviewer agreement regarding OM in the narrative, served as our reference standard classification for training and validation of machine learning models; trip reports without reviewer agreement were omitted (examples in Multimedia Appendix 1). We removed the blank space between words in all variables, except in narratives, to create single-text entities (ie, *DenverHealth* instead of *Denver Health*). We stemmed words and removed stop words (eg, *the*, *a*, or *and*). To prevent overfitting, an 80% training set and 20% test set were created. Training corpus was converted into a document term matrix (terms as columns and documents as rows) that described the frequency of terms that occurred in narratives. To classify trip reports (OM evidence: yes or no), we used NLP machine learning models available from the caret Package [18] on R version 3.4.1 (ie, random forest, k-nearest neighbors, support vector machines, and L1-regularized logistic regression). Values of hyperparameters and parameters for each model were estimated using default configurations (ie, no hyperparameter tuning), which were optimized with 3 repeats of 5-fold cross-validation and then fit to the entire training set. We assessed performance of each model by calculating PPV, negative predictive value (NPV), true-positive rates (TPRs), true-negative rates (TNRs), and areas under the receiver operating characteristic curves (AUCs), and we selected the binary classification algorithm with the highest AUC for subsequent model assessment. Details can be found in authored R code in Multimedia Appendix 2.

Phase 3: Validate Performance Measures Across Approaches

We searched for presence of the keywords *naloxone* (or *narcan*) among administered medications or *heroin* (or misspelled variations *herion* and *heroine*) in narratives of unseen September 2018 trip reports. Resulting trip reports were manually assessed following the same methodology as in phase 1. We then applied the machine learning classifier selected in phase 2 of the study to the reduced dataset of September 2018 trip reports. We hypothesized that machine learning models would decrease false-positive classifications of the combined *naloxone* and *heroin* search method because the algorithm would have learned and benefited from agreement in human assessments in phase 1. Reviewers' assessment was used as a reference standard to calculate PPV for each approach.

Results

Phase 1 Findings

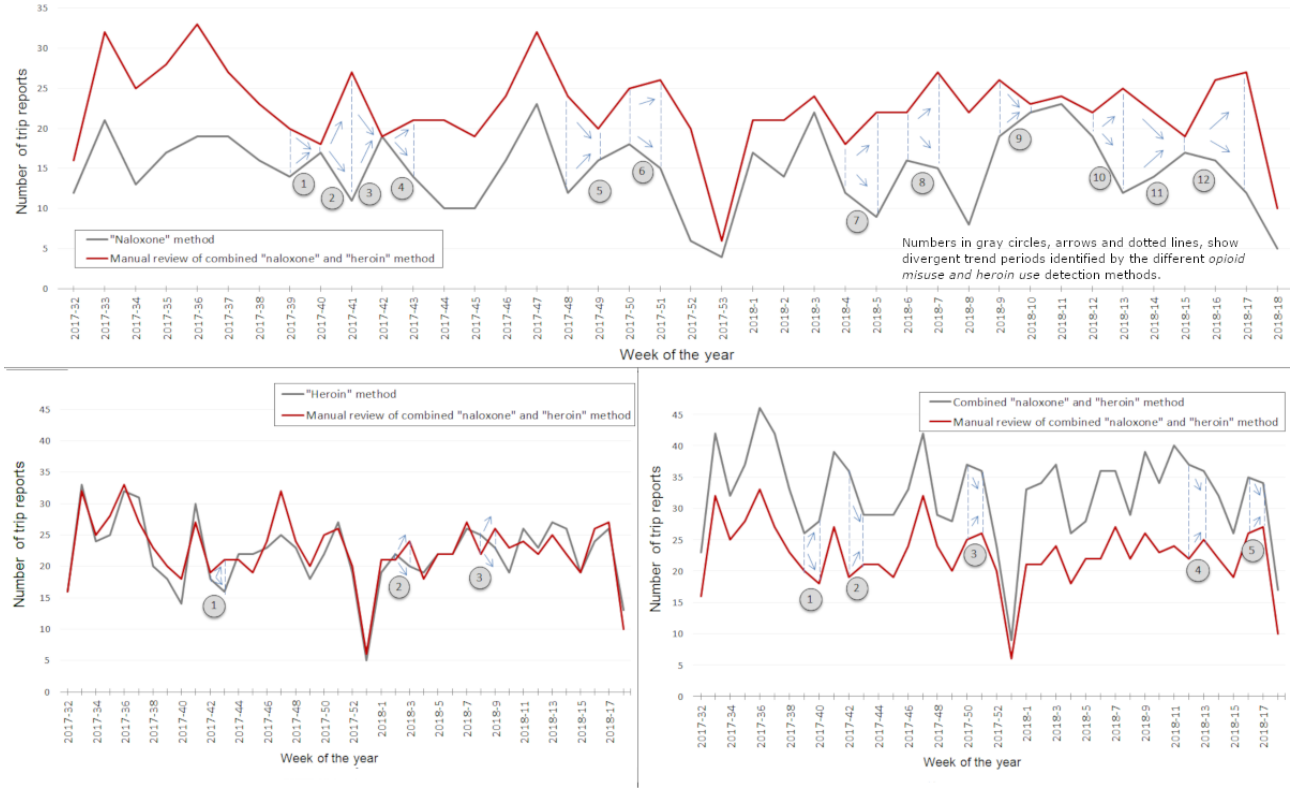
In total, 54,359 trip reports were filed, and 1.09% (594/54,359) indicated naloxone administration; reviewers agreed on assessment in 86.9% (516/594) of reports. Among trip reports with agreement, 56.6% (292/516) were considered to include information revealing OM.

Approximately 1.63% (884/54,359) of all trip reports mentioned *heroin* in the narrative. Reviewers agreed on potential OM assessment in 92.9% (821/884) of these. Among trip reports with agreement, almost all (784/821, 95.5%) were considered to include information revealing OM.

Combined results, where *naloxone* was administered by paramedics or *heroin* was mentioned in the narrative, accounted for 2.39% (1298/54,359) of trip reports. Reviewers agreed on potential OM assessment in trip reports in 89.83% (1166/1298) of these. Among trip reports with agreement, more than three-quarters (907/1166, 77.79%) included information consistent with OM.

Weekly counts of keywords mention varied by approach; Figure 1 is annotated to show periods of divergent trends between weekly sums of flagged reports and those affirmed by reviewer assessment. The *naloxone* approach was not consistent with reviewer assessment trends (r=0.60); the *heroin and* combined approaches were consistent with reviewer assessment trends (r=0.88 and r=0.90, respectively).

Figure 1. Weekly summary of paramedic trip reports trends for the documentation of administration of naloxone (top), heroin (bottom left), or both (bottom right), Denver Health, Denver, Colorado, August 1, 2017, to April 30, 2018.



Phase 2 Findings

The reference standard used to train and test machine learning models included details of 1166 *naloxone*- and *heroin*-flagged trip reports with positive OM reviewer assessment in phase 1. L1-regularized logistic regression was the highest performing algorithm (AUC=0.94; PPV=0.95; TPR=0.91; NPV=0.72; and TNR=0.84), followed by support vector machines (AUC=0.91; PPV=0.92; TPR=0.92; NPV=0.73; and TNR=0.73), random forest (AUC=0.91; PPV=0.91; TPR=0.95; NPV=0.79; and TNR=0.65), and k-nearest neighbors (AUC=0.81; PPV=0.79; TPR=1; NPV=0.1; and TNR=0.08). L1-regularized logistic regression yielded higher performance than the other algorithms; further statistical analyses, confusion matrices, and features that scored highest are presented in Multimedia Appendix 3.

Phase 3 Findings

Among 5983 September 2018 trip reports, *naloxone* identified 63 events, and chart review revealed 20 false positives (PPV=0.68). Examples of false positives are presented in Multimedia Appendix 4. Keyword *heroin* identified 129 trip reports, and chart review revealed 1 false positive (PPV=0.99). Combined *naloxone* and *heroin* searches identified 171 trip reports with 20 false positives (PPV=0.88).

L1-regularized logistic regression, the highest performing machine learning algorithm from phase 2, did not identify the one true negative of OM in reports flagged by *heroin* but identified 18 of the 20 true negatives of OM in reports flagged by naloxone administrations. The classifier identified 146 potential OM events from the 171 trip reports flagged by the combined text search with only 2 false positives. Results are summarized in Table 2. The machine learning classifier produced counts closer to those from reviewer assessment (Figure 2 shows counts for weeks 36 to 39 of 2018).

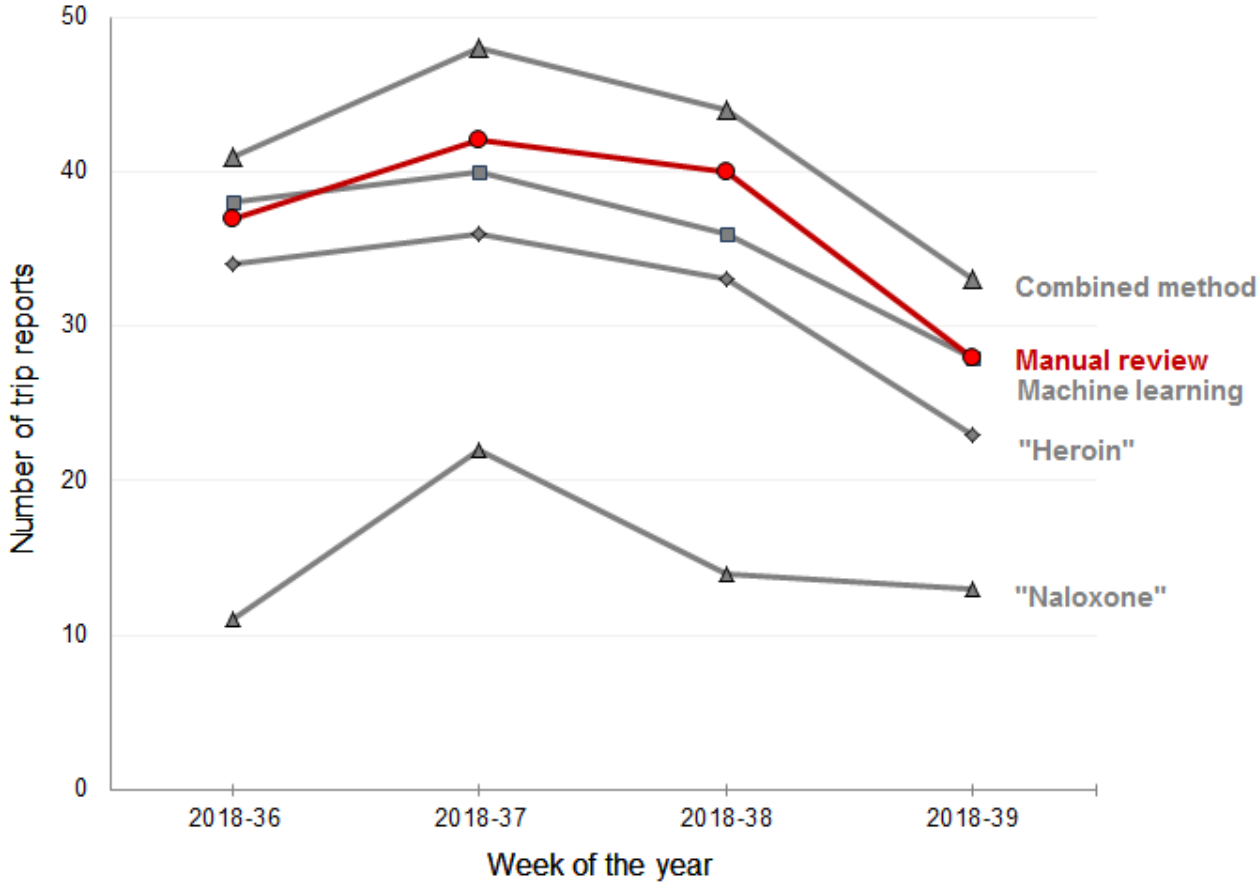
Table 2. Performance of natural language processing approaches to identify potential opioid misuse and heroin use in unseen September 2018 paramedic trip reports, Denver, Colorado.

Approach	Number of identified trip reports by approach (N)	Positive predictive value, n (%)	Correlation[a] between weekly opioid misuse counts and chart review assessment
Naloxone search among administered medications	63	43 (68.3)	0.86
Heroin search in narratives	129	128 (99.2)	0.99
Combined search approach (*naloxone* or *heroin*)	171	151 (88.3)	1
Machine learning[b] on combined search approach	146	144 (98.6)	0.99

[a]Pearson correlation coefficient.

[b]L1-regularized logistic regression.

Figure 2. Trends in weekly number of potential opioid misuse events by detection method in paramedic trip reports, Denver, Colorado, September 2018.



Discussion

Principal Findings

This study sought to better understand documentation in paramedic trip reports as a tool to support more effective nonfatal OM surveillance. Accurate detection of potential OM events in survivors of EMS runs can reflect short-term trends in OM-related events at the community and national levels. These are potential leading indicators for assessing the nonfatal magnitude of the opioid crisis in an area.

Fluctuating supplies and introduction of powerful, illicitly manufactured opioids may rapidly change local morbidity and mortality patterns [19,20]. Availability of near real-time data of opioid-related problems from the field may guide prevention and intervention efforts of emergency responders, health care providers, and public health practitioners [4]. Our methods, similar to those used to identify opioid overdose risk [21], could be applied to enhance information accuracy of EMS data for state and local public health departments, an important goal in the Centers for Disease Control and Prevention (CDC) Emergency Response Cooperative Agreement [22].

Public health agencies in the United States are seeking data sources and data-driven indicators for early warning systems to identify medical consequences of misuse of prescription and illicit opioids [23]. Our study found that naloxone administrations inaccurately identified and underestimated opioid-related paramedic trip events in Denver. This result is compatible with recent findings that naloxone administration was a poor proxy for opioid overdose [15]. Our study also found that EMS-administered naloxone did not reflect trends (rise or fall) in OM-related EMS runs assessed by chart review. By itself, EMS naloxone administration was a poor stand-alone indicator and would benefit from additional information embedded in EMS records.

As a simple alternative, the keyword *heroin* increased over 2.5-fold (from 63 flagged by the current standard [ie, naloxone administrations] to 171) the number of records with potential OM. This strategy flagged OM reports accurately, with only 1 false positive. Combined *naloxone* and *heroin* NLP search increased sensitivity but with substantial false positives. To improve this, we applied a machine learning algorithm that produced both higher sensitivity and specificity. This same tactic, previously employed to identify alcohol misuse in clinical notes of electronic health records [24], could be extended to include more opioid-related terms such as prescription opioid names. New studies should try to assess the effects of including records flagged by keywords such as *heroin* or opioid brand names in model training, testing, and validation.

Limitations

Two main limitations were present in this study. First, we used data from only 1 EMS system. Although DH paramedics adhere to a widely used data standard [17], implementation may vary between organizations. Second, calculation of the probability that cases not flagged by NLP methods were truly negative cases (NPV) was impossible as manual chart review of all trip reports would require human effort beyond our capacity.

Acknowledgments

JTP was hosted by Denver Public Health for his CDC Public Health Informatics Fellowship. The authors would like to thank Chad M Heilig (CDC), Scott H Lee (CDC), Matthew J Maenner (CDC), and Emily Bacon (DPH) for machine learning and statistics review and advice. The authors are grateful to the *Journal of Medical Internet Research* reviewers who made valuable suggestions to strengthen this manuscript and who highlighted new research avenues, which the authors hope to explore in forthcoming work. The authors of this study have no financial disclosures to report. This study is part of DH's Center for Addiction Medicine. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the CDC.

Authors' Contributions

JTP devised the study and led analysis, interpretation of data and results, and draft writing. KS contributed substantially to design and analysis. AJD contributed substantially to interpretation of data and results and draft writing. All authors contributed to interpretation of results and revision and approval of the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Excerpts of narratives in paramedic trip reports without reviewer agreement.
[DOCX File , 32 KB - jmir_v22i1e15645_app1.docx]

Multimedia Appendix 2
R code used in phase 2.
[TXT File , 6 KB - jmir_v22i1e15645_app2.txt]

Multimedia Appendix 3
Additional statistical analysis, confusion matrices, and feature scores by machine learning classifiers.
[DOCX File , 46 KB - jmir_v22i1e15645_app3.docx]

Multimedia Appendix 4
Excerpts of narratives of false positive results in phase 3.
[DOCX File , 32 KB - jmir_v22i1e15645_app4.docx]

References

1. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths — United States, 2013–2017. MMWR Morb Mortal Wkly Rep 2019;67(5152):1419-1427. [doi: 10.15585/mmwr.mm6751521e1]

2. Ahmad FB, Escobedo LA, Rossen LM, Spencer MR, Warner M, Sutton P. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics. 2019. Provisional Drug Overdose Death Counts URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm [accessed 2019-12-04]
3. Substance Abuse and Mental Health Services Administration. Substance Abuse and Mental Health Services Administration. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2018. Key Substance Use and Mental Health Indicators in the United States: Results From the 2018 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2019-11-26]
4. Vivolo-Kantor AM, Seth P, Gladden RM, Mattson CL, Baldwin GT, Kite-Powell A, et al. Vital signs: trends in emergency department visits for suspected opioid overdoses - United States, July 2016-September 2017. MMWR Morb Mortal Wkly Rep 2018 Mar 9;67(9):279-285 [FREE Full text] [doi: 10.15585/mmwr.mm6709e1] [Medline: 29518069]
5. Faul M, Dailey MW, Sugerman DE, Sasser SM, Levy B, Paulozzi LJ. Disparity in naloxone administration by emergency medical service providers and the burden of drug overdose in US rural communities. Am J Public Health 2015 Jul;105(Suppl 3):e26-e32. [doi: 10.2105/AJPH.2014.302520] [Medline: 25905856]
6. Knowlton A, Weir BW, Hazzard F, Olsen Y, McWilliams J, Fields J, et al. EMS runs for suspected opioid overdose: implications for surveillance and prevention. Prehosp Emerg Care 2013;17(3):317-329 [FREE Full text] [doi: 10.3109/10903127.2013.792888] [Medline: 23734988]
7. Ray BR, Lowder EM, Kivisto AJ, Phalen P, Gil H. EMS naloxone administration as non-fatal opioid overdose surveillance: 6-year outcomes in Marion County, Indiana. Addiction 2018 Dec;113(12):2271-2279. [doi: 10.1111/add.14426] [Medline: 30255531]
8. Merchant RC, Schwartzapfel BL, Wolf FA, Li W, Carlson L, Rich JD. Demographic, geographic, and temporal patterns of ambulance runs for suspected opiate overdose in Rhode Island, 1997-20021. Subst Use Misuse 2006;41(9):1209-1226. [doi: 10.1080/10826080600751898] [Medline: 16861173]
9. Lindstrom HA, Clemency BM, Snyder R, Consiglio JD, May PR, Moscati RM. Prehospital naloxone administration as a public health surveillance tool: a retrospective validation study. Prehosp Disaster Med 2015 Aug;30(4):385-389. [doi: 10.1017/S1049023X15004793] [Medline: 26061280]
10. Cash RE, Kinsman J, Crowe RP, Rivard MK, Faul M, Panchal AR. Naloxone administration frequency during emergency medical service events - United States, 2012-2016. MMWR Morb Mortal Wkly Rep 2018 Aug 10;67(31):850-853 [FREE Full text] [doi: 10.15585/mmwr.mm6731a2] [Medline: 30091966]
11. NIH National Institute on Drug Abuse. 2019. Opioid Overdose Crisis: Revised January 2019 URL: https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis [accessed 2019-12-02]
12. Hemsing N, Greaves L, Poole N, Schmidt R. Misuse of prescription opioid medication among women: a scoping review. Pain Res Manag 2016;2016:1754195 [FREE Full text] [doi: 10.1155/2016/1754195] [Medline: 27445597]
13. Pasero C, McCaffery M. Reversing respiratory depression with naloxone. Am J Nurs 2000 Feb;100(2):26. [Medline: 10683639]
14. Lewanowitsch T, Irvine RJ. Naloxone methiodide reverses opioid-induced respiratory depression and analgesia without withdrawal. Eur J Pharmacol 2002 Jun 7;445(1-2):61-67. [doi: 10.1016/s0014-2999(02)01715-6] [Medline: 12065195]
15. Grover JM, Alabdrabalnabi T, Patel MD, Bachman MW, Platts-Mills TF, Cabanas JG, et al. Measuring a crisis: questioning the use of naloxone administrations as a marker for opioid overdoses in a large US EMS system. Prehosp Emerg Care 2018;22(3):281-289. [doi: 10.1080/10903127.2017.1387628] [Medline: 29297739]
16. Gabow P, Eisert S, Wright R. Denver Health: a model for the integration of a public hospital and community health centers. Ann Intern Med 2003 Jan 21;138(2):143-149. [doi: 10.7326/0003-4819-138-2-200301210-00016] [Medline: 12529097]
17. National Emergency Medical Services Information System. National Emergency Medical Services Information System. Salt Lake City; 2016 Jul 13. NEMSIS Data Dictionary NHTSA v3.4.0 - EMS Data Standard URL: https://nemsis.org/media/nemsis_v3/release-3.4.0/DataDictionary/PDFHTML/DEMEMS/NEMSISDataDictionary.pdf [accessed 2019-11-26]
18. Kuhn M. CiteSeerX. Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien): The Comprehensive R Archive Network; 2015 Jul 16. A Short Introduction to the caret Package URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.696.4901&rep=rep1&type=pdf [accessed 2019-12-02]
19. Ciccarone D. Fentanyl in the US heroin supply: a rapidly changing risk environment. Int J Drug Policy 2017 Aug;46:107-111 [FREE Full text] [doi: 10.1016/j.drugpo.2017.06.010] [Medline: 28735776]
20. O'Donnell J, Gladden RM, Mattson CL, Kariisa M. Notes from the field: overdose deaths with carfentanil and other fentanyl analogs detected - 10 states, July 2016-June 2017. MMWR Morb Mortal Wkly Rep 2018 Jul 13;67(27):767-768 [FREE Full text] [doi: 10.15585/mmwr.mm6727a4] [Medline: 30001560]
21. Lo-Ciganic W, Huang JL, Zhang HH, Weiss JC, Wu Y, Kwoh CK, et al. Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. JAMA Netw Open 2019 Mar 1;2(3):e190968 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0968] [Medline: 30901048]
22. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. CDC-RFA-TP18-1802 Cooperative Agreement for Emergency Response: Centers for Disease Control and Prevention; 2018 Jun 20. CDC-RFA-TP18-1802 Cooperative Agreement for Emergency Response: Public Health Crisis Response 2018 Opioid Overdose Crisis Cooperative

Agreement Supplemental Guidance URL: https://www.cdc.gov/cpr/readiness/00_docs/TP18-1802OpioidSupplementalGuidance-508.pdf [accessed 2019-12-04]
23. Prieto JT, McEwen D, Davidson AJ, Al-Tayyib A, Gawenus L, Sangareddy SR, et al. Monitoring opioid addiction and treatment: do you know if your population is engaged? Drug Alcohol Depend 2019 Sep 1;202:56-60. [doi: 10.1016/j.drugalcdep.2019.07.002] [Medline: 31302412]
24. Afshar M, Phillips A, Karnik N, Mueller J, To D, Gonzalez R, et al. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. J Am Med Inform Assoc 2019 Mar 1;26(3):254-261. [doi: 10.1093/jamia/ocy166] [Medline: 30602031]

Abbreviations

AUC: area under the receiver operating curve
CDC: Centers for Disease Control and Prevention
DH: Denver Health
ED: emergency department
EMS: emergency medical services
NLP: natural language processing
NPV: negative predictive value
OM: opioid misuse
PPV: positive predictive value
TNR: true-negative rate
TPR: true-positive rate

Edited by CL Parra-Calderón; submitted 26.07.19; peer-reviewed by D Epstein, M Torii; comments to author 30.08.19; revised version received 05.09.19; accepted 08.10.19; published 03.01.20.

Please cite as:
Prieto JT, Scott K, McEwen D, Podewils LJ, Al-Tayyib A, Robinson J, Edwards D, Foldy S, Shlay JC, Davidson AJ
The Detection of Opioid Misuse and Heroin Use From Paramedic Response Documentation: Machine Learning for Improved Surveillance
J Med Internet Res 2020;22(1):e15645
URL: https://www.jmir.org/2020/1/e15645
doi:10.2196/15645
PMID:31899451

Original Paper

Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study

Ashley N D Meyer[1], PhD; Traber D Giardina[1], PhD; Christiane Spitzmueller[2], PhD; Umber Shahid[1], DrPH; Taylor M T Scott[1], BA; Hardeep Singh[1], MPH, MD

[1]Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, TX, United States

[2]Department of Psychology, University of Houston, Houston, TX, United States

Corresponding Author:
Ashley N D Meyer, PhD
Center for Innovations in Quality, Effectiveness and Safety
Michael E DeBakey Veterans Affairs Medical Center and Baylor College of Medicine
2002 Holcombe Blvd #152
Houston, TX
United States
Phone: 1 7134404660
Email: ameyer@bcm.edu

Abstract

Background: Patients are increasingly seeking Web-based symptom checkers to obtain diagnoses. However, little is known about the characteristics of the patients who use these resources, their rationale for use, and whether they find them accurate and useful.

Objective: The study aimed to examine patients' experiences using an artificial intelligence (AI)–assisted online symptom checker.

Methods: An online survey was administered between March 2, 2018, through March 15, 2018, to US users of the Isabel Symptom Checker within 6 months of their use. User characteristics, experiences of symptom checker use, experiences discussing results with physicians, and prior personal history of experiencing a diagnostic error were collected.

Results: A total of 329 usable responses was obtained. The mean respondent age was 48.0 (SD 16.7) years; most were women (230/304, 75.7%) and white (271/304, 89.1%). Patients most commonly used the symptom checker to better understand the causes of their symptoms (232/304, 76.3%), followed by for deciding whether to seek care (101/304, 33.2%) or where (eg, primary or urgent care: 63/304, 20.7%), obtaining medical advice without going to a doctor (48/304, 15.8%), and understanding their diagnoses better (39/304, 12.8%). Most patients reported receiving useful information for their health problems (274/304, 90.1%), with half reporting positive health effects (154/302, 51.0%). Most patients perceived it to be useful as a diagnostic tool (253/301, 84.1%), as a tool providing insights leading them closer to correct diagnoses (231/303, 76.2%), and reported they would use it again (278/304, 91.4%). Patients who discussed findings with their physicians (103/213, 48.4%) more often felt physicians were interested (42/103, 40.8%) than not interested in learning about the tool's results (24/103, 23.3%) and more often felt physicians were open (62/103, 60.2%) than not open (21/103, 20.4%) to discussing the results. Compared with patients who had not previously experienced diagnostic errors (missed or delayed diagnoses: 123/304, 40.5%), patients who had previously experienced diagnostic errors (181/304, 59.5%) were more likely to use the symptom checker to determine where they should seek care (15/123, 12.2% vs 48/181, 26.5%; *P*=.002), but they less often felt that physicians were interested in discussing the tool's results (20/34, 59% vs 22/69, 32%; *P*=.04).

Conclusions: Despite ongoing concerns about symptom checker accuracy, a large patient-user group perceived an AI-assisted symptom checker as useful for diagnosis. Formal validation studies evaluating symptom checker accuracy and effectiveness in real-world practice could provide additional useful information about their benefit.

(J Med Internet Res 2020;22(1):e14679) doi:10.2196/14679

KEYWORDS

clinical decision support systems; technology; diagnosis; patient safety; symptom checker; computer-assisted diagnosis

Introduction

Background

Patients are increasingly seeking to be more involved in their health care [1,2]. As a result, digital health care tools (both online and mobile health tools) have proliferated [3,4], and their use by patients has dramatically increased [5]. Overall, 1 in 3 US adults reported going online to try to self-diagnose a medical condition in 2013 [6]. In addition to searching the internet for health information, use of digital health care tools includes online, artificial intelligence (AI)–assisted symptom checkers for obtaining diagnoses or self-triage [7-10]. A previous report assessed the accuracy of general symptom checkers using patient vignettes [9] and found that diagnostic accuracy (defined as the correct diagnosis being listed first) was 34% and triage advice was appropriate 57% of the time. Accuracy varied considerably among symptom checkers (with a range of 5%-50%), leading to a concern about their use [11,12]. Furthermore, it is unknown if patients [7] use online symptom checkers as a replacement for seeing physicians in person. Also unknown are the rationale why patients use symptom checkers, whether they find them accurate and useful, and if these tools provide them with any benefit.

In light of evidence that approximately 1 in 20 US adults experience a diagnostic error every year (with half incurring severe or permanent harm) [13], the National Academies of Sciences, Engineering, and Medicine recommends the use of patient engagement tools, including symptom checkers and other digital health tools, in efforts to address this issue [14]. As a part of the solution digital health tools offer patients broader, quicker access to health information, [15], but their use may differ among patient groups. Mobile phone use for looking up general health information differs across race and ethnicity (with 67% of African Americans/blacks, 73% of Hispanics, and 58% of whites reportedly doing so) [16] and patients with chronic health conditions tend to have less access to the internet [17]. It is unclear how these patterns would relate to the use of online symptom checkers, but differences in use among these groups of patients could result in disparate benefits of the tools. Furthermore, other patient characteristics, such as previous positive or negative health care experiences, could also alter use, usefulness, and experiences with such tools.

Currently, it is unclear if patients use symptom checkers to supplement medical advice (which is what many of the tool developers suggest in addition to speaking with physicians about the obtained results) or if they are using them as a substitute for in-person health care by seeking in-person health care only if instructed by the symptom checker. Finally, in assessing symptom checker benefits, it is vital to understand patient perspectives [18] after actual use [19] (rather than to just assess their accuracy in fictitious situations as these data may not be ecologically valid). Knowledge about both the benefits of symptom checkers and how they can be improved could maximize patient benefits and minimize unintended consequences (such as *cyberchondria*, anxiety, or unnecessary health care use—proposed consequences of Web-based medical tools) [20-22].

Objectives

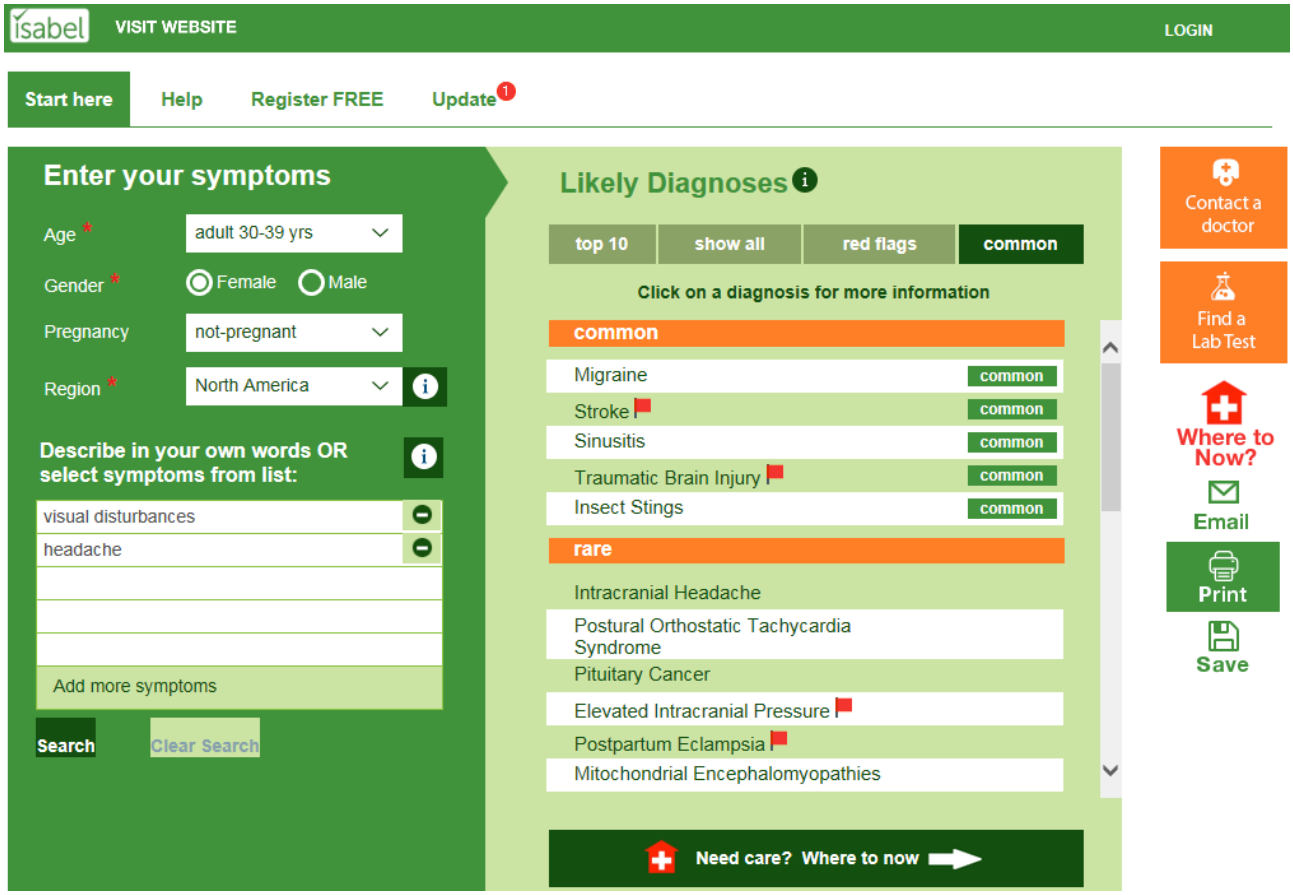
To address current knowledge gaps, we examined user characteristics and experiences and potential consequences of symptom checker use, including subsequent physician discussions around use of the symptom checker in relation to a popular online AI-assisted symptom checker, the Isabel Symptom Checker [23]. In addition, we compared perceptions of the symptom checker in patients who previously experienced errors in diagnosis versus those who did not, because these experiences may affect symptom checker favorability.

Methods

Description of the Isabel Symptom Checker

The Isabel Symptom Checker (Isabel Healthcare) [23] is a free Web-based, AI-assisted symptom checker intended for use by patients (as opposed to the Isabel Differential Diagnosis Generator [Isabel Healthcare] intended for clinicians) and has been shown to have better accuracy than the average symptom checker in a vignette-based study (defined as having the correct diagnosis listed first in 44% of cases compared with an average rate of 34% in the 23 symptom checkers tested) [9]. It currently has over 12,000 registered users globally, with almost 7000 in the United States (not all users register) and the symptom checker completes between 200,000 to 300,000 searches per month [24]. Patients research their symptoms by entering their age range, gender, pregnancy status, geographic location or travel history, and symptoms in everyday language. Using machine learning and a training database of 6000 disease presentations, the symptom checker uses evidence-based natural language processing techniques to create a list of likely diagnoses ranked in order of relevance for the symptoms entered. Patients can sort their likely diagnoses as a top-10 list; a full list of all relevant diagnoses; a list including only red-flag, *do-not-miss* diagnoses, which indicate that medical advice should be sought immediately; or as a list divided into common versus rare diagnoses. Diagnoses are linked to reference resources, allowing patients to learn more. These resources include the consumer-facing Merck Manual (Merck Sharpe & Dohme Corp) [25], MedlinePlus (National Library of Medicine) [26], a patient version of UpToDate (UpToDate, Inc) [27], and the Mayo Clinic website (Mayo Foundation for Medical Education and Research) [28]. Next steps are provided where users can "*contact a doctor*," "*find a lab test*," or determine where they should go for medical care using additional triage functionality (using the "*Where Now*" button). The symptom checker is freely available and provides information for both adult and pediatric patients (see Figure 1 for example screenshot).

Figure 1. Screenshot of the patient facing, artificial intelligence assisted Isabel Symptom Checker.



Participants

With the help of Isabel Healthcare, we sent email invitations to all registered US users of the Isabel Symptom Checker (4000) to complete an online survey through SurveyMonkey (SurveyMonkey) [29], a commercial survey website. All of these users had registered and used the symptom checker within the last 6 months. On the basis of the limited available internal institutional funding, we were able to offer a survey incentive to only the first 385 respondents, all of whom received US $20 gift card incentives; these available funds thus determined sample size. Local institutional review board approval was obtained at Baylor College of Medicine and written consent was obtained from the participants.

Survey

The survey was created by a multidisciplinary team (authors AM, TG, CS, and HS) with expertise in patient experience, cognitive psychology, psychometrics, internal medicine, and diagnostic errors. It comprises multiple-choice questions, 5-item Likert-type questions (with choices ranging from *strongly disagree* to *strongly agree*), and 5 open-ended questions and was designed to elicit information related to 4 main areas (see Multimedia Appendix 1 for full survey):

1. User characteristics (including age, gender, race, level of education, household income, and presence of chronic health conditions; see Multimedia Appendix 2 for full list).
2. Experiences of symptom checker use (including why and how patients used the tool, self-reported health and financial outcomes related to its use [in multiple-choice, Likert-type, and open-ended versions], whether they thought the symptom checker gave them useful information for their health problem, whether they followed the symptom checker's advice to go to the emergency department (ED) if advised to do so, how easy to use and useful the tool was, and whether the tool led them closer to correct diagnoses; see Multimedia Appendix 3 for full list).
3. Experiences discussing symptom checker results with physicians (whether patients discussed symptom checker use with physicians, and if so, physicians' receptiveness to patients' use of the tool [including an open-ended question further detailing those experiences], and if not, why they chose not to [in both multiple-choice and open-ended versions]; see Multimedia Appendix 4 for full list).
4. Personal experience of an error in diagnosis previously (defined for them as whether or not they have ever been given either the wrong diagnosis for a health concern or not given any diagnosis for a health concern that they were seeking medical help for; this includes both multiple-choice questions and an open-ended response, where participants could detail their diagnostic error experiences).

After development, the survey was pilot tested in both paper and online forms with 5 and 13 patients, respectively, and correspondingly refined to increase readability and understandability by simplifying and clarifying the language.

Data Analysis

All data were summarized using descriptive statistics, except open-ended responses, which were coded using content analysis. In addition, we compared demographics, experiences around Isabel Symptom Checker use, and subsequent interactions with physicians between users who had previously experienced diagnostic errors and those who had not using independent *t* tests, Chi-square, or Fisher exact tests where appropriate. We also conducted additional subanalyses using Chi-square or Fisher exact tests to determine whether certain behaviors (following the advice of the symptom checker and going to the ED and talking to one's doctor about Isabel results) were associated with other demographics. All tests were two tailed, done using IBM SPSS Statistics 22 (IBM Corporation), and considered significant when $P<.05$.

Results

Sample

From the sample of 385 respondents, 329 provided mostly complete (>90% of the survey was complete) and relevant data (18 participants' responses were excluded for not completing >90% of the questions and 38 because they described using the tool as a medical professional for either education or diagnosing patients when elaborating on the question "What prompted you to use the Isabel Symptom Checker?" after choosing the "Other" response). Only data from the 329 nonexcluded respondents are reported. The mean time to complete the survey was 12:21 (SD 10:43) min.

User Characteristics

Mean respondent age was 48.0 (SD 16.7) years; most of them were women (230/304, 75.7%), white (271/304, 89.1%), with bachelor's degrees or higher (191/302, 63.2%), and had less than US $100,000 in household income (216/287, 75.3%), health care coverage (296/304, 97.4%), and chronic health conditions (216/329, 65.7%; Multimedia Appendix 2).

Experiences Around Symptom Checker Use

Patients most commonly used the symptom checker to better understand what could cause their symptoms (232/304, 76.3%). The next most common reasons included to decide whether to seek in-person health care (101/304, 33.2%), to decide what health care setting to visit (eg, primary or urgent care: 63/304, 20.7%), to get medical advice without going to the doctor (48/304, 15.8%), or to better understand the diagnosis made by their doctor (39/304, 12.8%). Many respondents used the symptom checker before (119/304, 39.1%) or both before *and* after seeing a physician (113/304, 37.2%). Of additional note, of the 26 patients given advice to proceed to the ED, 14 (54%) did. Most users thought the symptom checker gave them useful information for their health problems (274/304, 90.1% either strongly agreeing or agreeing) with about half reporting positive health effects (154/302, 51.0%). Although over half were neutral in terms of benefitting financially (172/303, 56.8%); most found the symptom checker useful: perceiving it to be satisfying (263/304, 86.5%), easy to use (182/303, 60.1%), useful as a diagnostic tool (253/301, 84.1%), and providing them with insights leading them closer to correct diagnoses (231/303, 76.2%). In addition, most reported they would use it again (278/304, 91.4%; Multimedia Appendix 3).

Open-ended responses detailing effects on health based on what participants learned most often included positive consequences (168/175, 96.0%). Most often, patients conveyed that symptom checker use enabled them to determine whether their condition might be serious, which helped them distinguish when to seek medical attention based on symptoms and severity (49/175, 28.0%; see Multimedia Appendix 5 for additional findings). Similarly, open-ended responses about financial effects were mostly positive (64/69, 93%) and most often related to reporting fewer doctor visits post–symptom checker use (34/69, 49%; see Multimedia Appendix 5 for additional findings).

Experiences Discussing Symptom Checker Results With Physicians

Of those who visited physicians after using the tool (213/304, 70.1%), almost half discussed the findings with their physicians (103/213, 48.4%). Their experiences were mixed, but patients more often felt physicians were interested (42/103, 40.8%) than *not* interested (24/103, 23.3%) in learning about the tool's results. Similarly, patients more often felt their physicians were open (62/103, 60.2%) than *not* open (21/103, 20.4%) to discussing the tool's results (Multimedia Appendix 4). In open-ended responses, patients described both positive (15/29, 52%) and negative (14/29, 48%) interactions with their doctors when discussing their Isabel results. For example, the most often talked about positive experience discussing the results with physicians was the perception that physicians were open to the use of Isabel (6/29, 21%), yet the most often talked about negative experience was frustration on behalf of the patients during such discussions (7/29, 24%; see Multimedia Appendix 5 for additional findings).

Patients who chose *not* to discuss the findings with their physicians (110/213, 51.6%) did so because of various concerns, including thinking their doctors would not approve of their use of the tool or the doctors would think the patients mistrusted them or were trying to second guess or replace them by using the tool (see Multimedia Appendix 4). In the corresponding open-ended response, they most often described not discussing the results with their doctors because of worry about pushback or concerns about their physicians' reactions (21/52, 40%; see Multimedia Appendix 5 for additional findings).

Previous Experiences of Diagnostic Errors

More than half of the patients reported previously experiencing diagnostic errors (181/304, 59.5%). Females made up 80.7% (146/181) of the diagnostic error group but only 68.3% (84/123) of the nonerror group (see Multimedia Appendix 2; $P=.01$). In addition, patient users reporting previous diagnostic errors reported having more doctor visits in the last year (10.4 vs 4.1 visits; $P<.001$); had higher use of online resources to obtain medical information, including sources other than WebMD or Google (35/181, 19.3% vs 10/123, 8.1%; $P=.01$); and were more likely to have arthritis (88/169, 52.1% vs 38/121, 31.4%; $P<.001$), asthma (47/166, 28.3% vs 15/117, 12.8%; $P=.002$), or other chronic health conditions (93/181, 51.4% vs 25/123,

20.3%; *P*<.001) compared with the nonerror group (Multimedia Appendix 2).

Users who previously experienced diagnostic errors were also more likely to use the Isabel Symptom Checker to determine where they should seek care (48/181, 26.5% vs 15/123, 12.2%; *P*=.002) and to use it both before and after seeing a doctor (rather than at only one time point; 86/181, 47.5% vs 27/123, 22.0%; *P*<.001). They were also more likely to experience positive health benefits from the symptom checker compared with others (98/181, 54.1% vs 56/121, 46.3%; *P*=.03). The diagnostic error group was also more likely to perceive they obtained insights about their diagnoses from the tool (141/180, 78.3% vs 90/123, 73.2%; *P*=.01), and less often found their doctors supportive regarding their use of the tool (61/179, 34.1% vs 48/121, 39.7%; *P*=.02; Multimedia Appendix 3).

Users who previously experienced diagnostic errors were more likely to see a doctor after using the symptom checker than those who did not (145/181, 80.1% vs 68/123, 55.3%; *P*<.001), but they were equally likely to discuss the results with their physicians (69/145, 47.6% vs 34/68, 50.0%; *P*=.74). In these conversations, however, they less often felt their doctors were interested in learning about their symptom checker results (22/69, 32% vs 20/34, 59%; *P*=.04; Multimedia Appendix 4).

When describing their diagnostic errors in open-ended responses (n=108), patients reported several contributory factors to their diagnostic errors. These included their perceptions that physicians (1) were unable to manage diagnostic uncertainty (33/108, 30.6%), (2) made multiple unnecessary referrals to others when faced with challenging diagnoses (22/108, 20.4%), (3) prioritized financial gains over patient benefit (19/108, 17.6%), (4) unfairly labelled patients (eg, as drug or attention seekers, as *drama queens*, as having symptoms “all in [their] head[s],” or as not “look[ing] sick”: 16/108, 14.8%), and (5) did not take the time to listen to patients (13/108, 12.0%). Several patients reported harm, including long-term health consequences from errors, such as disability or life-threatening experiences (73/108, 67.6%; see Multimedia Appendix 5 for additional findings).

Additional Behavioral Differences as Related to Demographics

Neither likelihood of going to the ED when the symptom checker suggested (n=25) nor the likelihood of discussing the results with their doctors (assuming they saw a doctor after using the symptom checker; n=217) were significantly related to gender, income, education, or being an underrepresented minority in our sample (see Multimedia Appendix 5 for details).

Discussion

Principal Findings

Patients used an online symptom checker to learn more about what could cause their symptoms, to determine whether to seek care or where, to get medical advice without going to a doctor, or to better understand their diagnosis. Most patients thought the tool gave them useful information for their health problems and thought it provided them with insights leading them closer to correct diagnoses. Half of the patients reported positive health effects. However, the patients who discussed the findings with their physicians conveyed mixed experiences about whether physicians were interested or open about discussing symptom checker results.

Strengths

The strengths of this study are the examination of naturalistic patient experiences and the assessment of subsequent related events, which are often missing from existing digital health tool studies (most previous studies examined vignette-based assessments [30,31] or patients already presenting to their doctors [32-35] with limited follow-up) [7]. Most patients used the symptom checker between 2 weeks and 4 months before the survey, allowing for adequate time for diagnoses to evolve and related subsequent events to occur, such as the completion of diagnostic tests, referrals, treatment, and potential responses to treatment.

Limitations

However, there are several study limitations. As we rely on self-reported data, there is no validation of patient outcomes via some type of medical record audit, making it difficult to assess outcome accuracy. Nonetheless, over time, patients would have enough information to make a determination about the ultimate accuracy of the diagnosis suggested by the tool. In addition, as with all surveys, participants may be subject to acquiescence bias—the tendency to agree with most statements. However, we did not find much evidence for this: despite much agreement with positively worded questions, negatively worded questions were not similarly agreed with (people were not merely agreeing). An additional limitation is that these data represent patient perceptions of only 1 symptom checker, and it is not clear if these results would generalize to other symptom checkers, especially to those that do not utilize AI-assisted natural language searching. We also offered an incentive of a US $20 gift card to the first 385 participants, which may have skewed our sample to people who are quick to respond to emails. Our sample might also be unique: participants had a mean of 8 visits to physicians within the last 12 months, meaning they could be different—perhaps sicker—compared with the general population. However, this population may also be more likely to use such tools given their high interaction with the health care system, so these patterns are still important to understand. In addition, our sample is overwhelmingly female and white, with a mean age of 48 years, thereby reducing our ability to examine demographic differences in terms of experiences or behavior related to symptom checker use. However, this represents user data available from Isabel Healthcare (females represented 62% of users over the last year, with 39% of users aged between 40 and 64 years). It is difficult to know if our sample is representative of typical users in other ways. Finally, this study was not designed to explain the differences in perceptions and experiences between groups who had experienced diagnostic errors versus those who had not, but only to describe them: the reasons for these differences are likely very complicated and future studies could further examine the roots of these differences.

For Additional Discussion

Some findings warrant additional consideration. For example, previous studies show that some underrepresented groups use mobile resources more for obtaining health information [16]. Perhaps these groups are using digital health tools as a substitute for other less-available health resources. Given that the long-term implications of using these tools are not understood, this could represent disparities affecting health outcomes, especially as patients in this study used the tool to triage themselves or get medical advice without going to a doctor. Nonetheless, our sample did not overwhelmingly include underrepresented groups. As such, additional research is needed to further scrutinize disparities related to symptom checker use.

Another finding worth additional consideration is that over half of the respondents reported previously experiencing diagnostic errors. Although this may seem high, this is a selected sample of symptom checker users, many of whom have had multiple interactions with the health care system. We do not intend this to be a population-based estimate. Nonetheless, the National Academies of Sciences, Engineering, and Medicine have extrapolated from large estimates that most Americans will get a wrong or late diagnosis at some point in their lives [14], and population-based surveys suggest that 12% of patients may have been misdiagnosed, so the high rate of misdiagnosis is quite possible in our sample [36]. These patients used the tool at more time points and used more online health resources in general, but they perceived their doctors to be less interested when discussing the tool's results. This could relate to the higher incidence of chronic diseases reported in this group and more negative health care experiences that often occur in patients with chronic disease [37]. Although past dissatisfaction with the health care system has been linked to increased use of the internet for health-related purposes [38-40], the impact of medical circumstances or past diagnostic errors on the use of alternate health resources (such as symptom checkers) remains ripe for exploration.

Our findings also highlight a disconnect between patients and physicians when it comes to the use of digital health tools. Although the sample was generally enthusiastic and satisfied with the tool, the patients felt their physicians showed mixed receptivity to the information and mixed openness to discussing it. This might discourage future use of such tools and future engagement by patients, similar to patterns seen in the contrasting patient and physician enthusiasm about email use for health communications [41].

In addition to this concern, a fear that has surfaced over the use of these tools is the potential for patients' anxiety to increase, thereby increasing health care utilization. These data show that many patients are using the tool to see whether they needed to see a doctor and help them determine where they should seek care. Despite this, a previous study pointed out that this particular symptom checker never advises self-care, which may also increase health care utilization [9]. We currently do not know if such tools would lead to a significant increase in health care use. A larger sample and additional objective follow-up data would help us understand if this represents appropriate utilization of resources.

Finally, we think it is worth reflecting on the effect that such tools might have on patients' sense of confidence in their abilities to diagnose themselves. Diagnosis is a task that often involves clinical uncertainty, something physicians themselves face [42]. Undoubtedly, patients would experience more diagnostic uncertainty than physicians owing to less expertise, but as more patients use these types of tools and obtain answers without actually seeing a health care professional, it will be important to examine the effect of these tools on how patients think about self-diagnosis and any resulting consequences thereof (such as false reassurance, suggested by others [43]). This study is an initial examination of real-life symptom checker use, but as Fraser et al point out [43], the evaluation of such tools should assess them with increasing ecological validity and should examine multiple aspects: usability, effectiveness, and safety. We have begun to examine usability and effectiveness, but much more remains to be understood to thoroughly investigate all of these facets in real-world situations.

Conclusions

In conclusion, while accessing a popular online symptom checker for triage and diagnosis, patients reported receiving useful information for their diagnostic process, despite ongoing concerns about the accuracy of various types of symptom checkers [43]. Prior negative health care experiences related to misdiagnoses might affect how patients use and benefit from these tools for triage and diagnosis, an area ripe for exploration. Evaluation of long-term, objective health benefits, particularly in diverse patient groups, is needed to better understand the broader impact of symptom checkers on diagnosis and health outcomes.

Acknowledgments

The authors would like to thank Jason Maude, founder and CEO of Isabel Healthcare, for distributing the survey to users and for providing product information regarding the symptom checker. This work was supported in part by the Houston Veterans Administration (VA) Health Services Research and Development (HSR&D) Center for Innovations in Quality, Effectiveness, and Safety (CIN13-413). In addition, AM is supported by a VA HSR&D Career Development Award (CDA-17-167); TG is supported by an Agency for Healthcare Research and Quality Mentored Career Development Award (KO1-HS025474); and HS is supported by the VA HSR&D Service (CRE12-033 and the Presidential Early Career Award for Scientists and Engineers USA 14-274), the VA National Center for Patient Safety, the Agency for Healthcare Research and Quality (R01HS27363), and the Gordon and Betty Moore Foundation. These funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to report. Jason Maude, founder and CEO of Isabel Healthcare, distributed the survey to users and provided the product information regarding the symptom checker, but otherwise he did not have input on the analysis, conclusions reached, or manuscript preparation and did not commission this report or provide funding for it.

Multimedia Appendix 1
Full survey (delivered via SurveyMonkey): Isabel Symptom Checker Post-Use Survey.
[DOCX File , 27 KB - jmir_v22i1e14679_app1.docx]

Multimedia Appendix 2
Characteristics of the Isabel Symptom Checker patient users.
[DOCX File , 24 KB - jmir_v22i1e14679_app2.docx]

Multimedia Appendix 3
Patient experiences of using the Isabel Symptom Checker.
[DOCX File , 27 KB - jmir_v22i1e14679_app3.docx]

Multimedia Appendix 4
Patients' experiences of discussing the Isabel Symptom Checker results with their physicians.
[DOCX File , 24 KB - jmir_v22i1e14679_app4.docx]

Multimedia Appendix 5
Additional findings from open-ended responses.
[DOCX File , 24 KB - jmir_v22i1e14679_app5.docx]

References

1. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. BMJ Qual Saf 2013 Oct;22(Suppl 2):ii33-ii39 [FREE Full text] [doi: 10.1136/bmjqs-2012-001623] [Medline: 23893394]
2. Anderson M, McCleary KK. From passengers to co-pilots: patient roles expand. Sci Transl Med 2015 Jun 10;7(291):291fs25. [doi: 10.1126/scitranslmed.aac6023] [Medline: 26062844]
3. Ralf-Gordon J. Research 2 Guidance.: Research 2 Guidance; 2016. The mHealth Apps Market is Getting Crowded URL: https://research2guidance.com/mhealth-app-market-getting-crowded-259000-mhealth-apps-now/ [accessed 2019-12-06]
4. Aitken M, Lyle J. IQVIA. Parsippany, NY: IMS Institute for Healthcare Informatics; 2015 Sep. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2019-12-06]
5. Powell AC, Landman AB, Bates DW. In search of a few good apps. J Am Med Assoc 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
6. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013-2/ [accessed 2019-12-06]
7. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018 Sep 25;5(3):95-105. [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
8. Jutel A, Lupton D. Digitizing diagnosis: a review of mobile applications in the diagnostic process. Diagnosis (Berl) 2015 Jun 1;2(2):89-96. [doi: 10.1515/dx-2014-0068] [Medline: 29540025]
9. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015 Jul 8;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
10. Wyatt JC. Fifty million people use computerised self triage. Br Med J 2015 Jul 8;351:h3727. [doi: 10.1136/bmj.h3727] [Medline: 26156750]
11. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. Phys Med Rehabil 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
12. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. Int J Bipolar Disord 2017 Dec;5(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0073-9] [Medline: 28155206]
13. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]

14. National Academies of Sciences, Engineering, and Medicine, Institute of Medicine, Board on Health Care Services, Committee on Diagnostic Error in Health Care. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015.
15. Akter S, Ray P. mHealth - an ultimate platform to serve the unserved. Yearb Med Inform 2010;2010:94-100. [Medline: 20938579]
16. Anderson M. Pew Research Center. 2015 Apr 30. Racial and Ethnic Differences in How People Use Mobile Technology URL: http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2019-12-06]
17. Fox S, Duggan M. Pew Research Center. 2013 Nov 26. The Diagnosis Difference URL: https://www.pewresearch.org/science/2013/11/26/the-diagnosis-difference/ [accessed 2019-12-06]
18. Vincent CA, Coulter A. Patient safety: what about the patient? Qual Saf Health Care 2002 Mar;11(1):76-80 [FREE Full text] [doi: 10.1136/qhc.11.1.76] [Medline: 12078376]
19. McCartney M. Margaret McCartney: innovation without sufficient evidence is a disservice to all. Br Med J 2017 Sep 5;358:j3980. [doi: 10.1136/bmj.j3980] [Medline: 28874353]
20. White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. ACM Trans Inf Syst 2009;27(4):1-37. [doi: 10.1145/1629096.1629101]
21. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
22. Gass MA. Digital Commons at Salem State University. Salem State University: Department of Nursing, Salem State University; 2016 May 1. Risks and Benefits of Self-diagnosis Using the Internet URL: https://digitalcommons.salemstate.edu/cgi/viewcontent.cgi?article=1096&context=honors_theses [accessed 2019-12-06]
23. Isabel Symptom Checker. 2018. Isabel Symptom Checker - The One the Doctors Use URL: https://symptomchecker.isabelhealthcare.com/the-symptom-checker/how-it-works [accessed 2019-12-06]
24. Maude J. Meyer AN. 2018. Personal email communication .
25. Merck Manuals. 2019. Merck Manual Consumer Version URL: https://www.merckmanuals.com/home [accessed 2019-12-06]
26. MedlinePlus. 2019. URL: https://medlineplus.gov/ [accessed 2019-12-06]
27. UpToDate. 2019. UpToDate Subscriptions for Patients and Caregivers URL: https://www.uptodate.com/home/uptodate-subscription-options-patients [accessed 2019-12-06]
28. Mayo Foundation for Medical Education Research (MFMER). Mayo Clinic. 2019. URL: https://www.mayoclinic.org/ [accessed 2019-12-06]
29. SurveyMonkey. 2019. URL: https://www.surveymonkey.com/ [accessed 2019-12-06]
30. Coiera E. The Guide to Health Informatics 3rd Edition. 2018 Jun 29. Paper Review: The Babylon Chatbot URL: https://coiera.com/2018/06/29/paper-review-the-babylon-chatbot/ [accessed 2019-12-06]
31. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
32. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Aug 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
33. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. How accurate are patients at diagnosing the cause of their knee pain with the help of a web-based symptom checker? Orthop J Sports Med 2016 Feb;4(2):2325967116630286 [FREE Full text] [doi: 10.1177/2325967116630286] [Medline: 26962542]
34. Hageman MG, Anderson J, Blok R, Bossen JK, Ring D. Internet self-diagnosis in hand surgery. Hand (N Y) 2015 Sep;10(3):565-569 [FREE Full text] [doi: 10.1007/s11552-014-9707-x] [Medline: 26330798]
35. Farmer SE, Bernardotto M, Singh V. How good is internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? Clin Otolaryngol 2011 Oct;36(5):517-518. [doi: 10.1111/j.1749-4486.2011.02375.x] [Medline: 22032458]
36. NORC at the University of Chicago, IHI/NPSF Lucian Leape Institute. Americans' Experiences with Medical Errors and Views on Patient Safety. In: Institute for Healthcare Improvement. Cambridge, MA: Institute for Healthcare Improvement and NORC at the University of Chicago; 2017.
37. Ali A, Vitulano L, Lee R, Weiss TR, Colson ER. Experiences of patients identifying with chronic Lyme disease in the healthcare system: a qualitative study. BMC Fam Pract 2014 May 1;15:79 [FREE Full text] [doi: 10.1186/1471-2296-15-79] [Medline: 24885888]
38. de Rosis S, Barsanti S. Patient satisfaction, e-health and the evolution of the patient-general practitioner relationship: evidence from an Italian survey. Health Policy 2016 Nov;120(11):1279-1292 [FREE Full text] [doi: 10.1016/j.healthpol.2016.09.012] [Medline: 27836231]
39. Li N, Orrange S, Kravitz RL, Bell RA. Reasons for and predictors of patients' online health information seeking following a medical appointment. Fam Pract 2014 Oct;31(5):550-556. [doi: 10.1093/fampra/cmu034] [Medline: 24963151]
40. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]

41. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009 Jun 16;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]
42. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H. Defining and measuring diagnostic uncertainty in medicine: a systematic review. J Gen Intern Med 2018 Jan;33(1):103-115 [FREE Full text] [doi: 10.1007/s11606-017-4164-1] [Medline: 28936618]
43. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Nov 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]

Abbreviations

AI: artificial intelligence
ED: emergency department
HSR&D: Health Services Research and Development
VA: Department of Veterans Affairs

Edited by G Eysenbach; submitted 12.05.19; peer-reviewed by K McDonald, D Wong, N Mackintosh, S D'Arcy; comments to author 22.08.19; revised version received 17.10.19; accepted 22.10.19; published 30.01.20.

Please cite as:
Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H
Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study
J Med Internet Res 2020;22(1):e14679
URL: http://www.jmir.org/2020/1/e14679/
doi:10.2196/14679
PMID:32012052

Original Paper

Health Care Personnel's Perspective on Potential Electronic Health Interventions to Prevent Hospitalizations for Older Persons Receiving Community Care: Qualitative Study

Martha Therese Gjestsen[1], RN, MSc, PhD; Siri Wiig[2], MSc, PhD; Ingelin Testad[1,3], RN, MSc, PhD

[1]Centre for Age-related Medicine, Stavanger University Hospital, Stavanger, Norway

[2]Centre for Resilience in Healthcare, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

[3]Exeter University Medical School, Exeter University, Exeter, United Kingdom

Corresponding Author:
Martha Therese Gjestsen, RN, MSc, PhD
Centre for Age-related Medicine
Stavanger University Hospital
Jan Johnsensgate 12
Stavanger, 4011
Norway
Phone: 47 92805525
Email: gjemja@sus.no

Abstract

Background: The use of electronic health (eHealth) interventions is suggested to help monitor and treat degenerative and chronic diseases through the use of sensors, alarms, and reminders and can potentially prevent hospitalizations for home-dwelling older persons receiving community care. It is increasingly recognized that the health care personnel's acceptance of a technological application remains a key challenge in adopting an intervention, thus interventions must be perceived to be useful and fit for purpose by the actual users.

Objective: The aim of this study was to identify and explore the perspectives of managers and health care personnel in community care regarding the use of eHealth interventions in terms of prevention of hospitalizations for home-dwelling older persons receiving community care.

Methods: A case study with a qualitative approach was carried out in community care in a Norwegian municipality, comprising individual interviews and focus group interviews. A total of 5 individual interviews and 2 focus group interviews (n=12) were undertaken to provide the health care personnel's and managers' perspective regarding the use of eHealth interventions, which could potentially prevent hospitalizations for home-dwelling older persons receiving community care. Data were analyzed by way of systematic text condensation, as described by Malterud.

Results: The data analysis of focus group interviews and individual interviews resulted in 2 categories: potential technological applications and potential patient groups. Discussions in the focus groups generated several suggestions and wishes related to technical applications that they could make use of in their day-to-day practice. The health care personnel warranted tools and measures to enhance and document their clinical observations in contact with patients. They also identified patient groups, such as patients with chronic obstructive pulmonary disease or dehydration or urinary tract infections, for whom hospitalizations could potentially have been prevented.

Conclusions: We have shown that the health care personnel in community care warrant various technological applications that have the potential to improve quality of care and resource utilization in the studied municipality. We have identified needs and important matters in practice, which are paramount for acceptance and adoption of an intervention in community care.

(J Med Internet Res 2020;22(1):e12797) doi:10.2196/12797

KEYWORDS

health services research; community health services; hospitalization; health services for the aged; qualitative research; focus groups; eHealth; technology

Introduction

Background

The global shift in demographics represents an epidemiological transition from a predominance of infectious diseases to noncommunicable diseases (ischemic heart disease, stroke, and chronic lung disease) [1]. The use of electronic health (eHealth) is suggested to help monitor and treat degenerative and chronic diseases through the use of sensors, alarms, and reminders [2-5].

The underlying assumption is that the use of digital technologies can potentially redesign care pathways in a way that will improve monitoring and treatment of degenerative and chronic diseases, encourage better self-management of health problems, and alert professional support if devices signal a problem [2,3,6-9], ultimately reducing the disruptive impact of acute unscheduled hospital admissions, for example, for older persons [10]. Previous research has identified that emergency hospital admissions often occur when an older person has reached a point of crisis because of a combination of circumstances, such as an exacerbation of a chronic condition, change in social setting, or a cascade of symptoms because of multimorbidity and frailty [11-17]. The use of eHealth could thus be applied as a tool to prevent a severe state of illness that requires hospitalization by discovering and addressing the patients' symptoms at an early stage.

In addition, from a resource perspective, the prevention of hospitalizations for older persons has gained much attention in the last decade [14,15,18,19]. Persons aged older than 65 years are substantial consumers of hospital care; there is a peak in hospitalization rates for both men and women in the age group of 80 years and older in all European countries [20]. Increasing age is thus associated with an increasing demand for specialized health care [21-23], and this may threaten the sustainability of the health care systems, as a larger share of older persons in the population implies a dwindling proportion of the workforce, consequently likely to aggravate existing strains on formal health systems [24,25]. Norwegian policy documents emphasize that a major response to the resource challenge in health care is to enable and empower people to live in their own home for as long as possible, as well as provide timely treatment interventions at the proper level in the health care system. The government introduced the *Coordination Reform January 1, 2012*, which represents a transition in responsibility for providing health care services, where the municipalities are to play a much larger role in meeting the demand for services [26]. In the reform, the preventative perspective in health care is of great focus, where an important assumption is that there is a potential for preventing hospital admissions for the older persons receiving community care.

Despite the rhetoric associated with the benefits of adopting eHealth interventions in community care to prevent hospitalizations, the use of such technologies has not developed at the pace and scale anticipated [27]. The resource and safety challenges are appropriate and well-rehearsed incentives to adopt certain technology interventions, but it is increasingly recognized in the research literature that the health care personnel's acceptance of the technological application itself remains a key challenge in adopting an intervention [28-30], underlining the vital importance that the involved stakeholders (eg, researchers, policy makers, health care personnel, patients, and carers) are able to judge the value of an eHealth intervention in its own right. Conversely, until we develop interventions that are considered to be useful and fit for purpose by the actual users, there will be reluctance regarding adoption of technologies in health care [30,31].

A review by Joseph et al [32] found that identifying issues and needs in practice were the main challenges related to the development and implementation of telehealth projects. This implies that identification of patients who might benefit from an intervention and a clearly defined role of a technological application (whether it is a new application, a new clinical tool, or a new system for delivering care remotely) are factors paramount for acceptance and adoption of an intervention [27,33]. These aspects are, however, not described in the body of research concerning the development of eHealth interventions in community care. Consequently, knowledge concerning the health care personnel and managers in this context of care is scarce.

On the basis of the notion that the managers and health care personnel in community care play a pivotal role in informing an eHealth intervention, it is of vital importance to explore their perspectives, thus gaining a better understanding of which technology-based interventions are deemed to be more appropriate and which patient groups an intervention could target. Ultimately, this knowledge can contribute to a more optimized intervention by increasing the probability for staff acceptance, as the intervention is developed on the basis of needs and suggestions defined by the managers and health care personnel in community care.

Informing an Electronic Health Intervention in Community Care

EHealth interventions are suggested as a means to improve efficiency, quality, and safety of care [33,34]. According to previous research, the adoption and implementation of such interventions in a complex health care system is challenging. In complex systems, elements are interdependent and mutually reinforcing; they interact with other systems in unexpected ways, as they comprise several aspects, such as technologies, humans and its social environment, which can simultaneously be members of several interrelated systems [35,36]. This sociotechnical perspective recognizes that people, technologies, organizations, and process of care interact in complex ways [37-39]. The unique competence that the nurses are in possession of should be taken into account to optimize the uptake and use of an eHealth intervention. By nature, the intervention is intimately and reciprocally entwined with the professional skills and networks that support technology use and the development of community care services and with the local, national, and transnational policy on technological innovation and assisted living [37,40,41]. Thus, nurses who provide care using eHealth must be well-grounded in general nursing knowledge, theory, and practice competencies and should furthermore have clinical experience and capacity to possess attributes of intuition and creativity to enhance a holistic care [42].

On this backdrop, this paper focusses on building a rationale for adopting an eHealth intervention in community care by exploring the perspectives provided by the health care personnel and managers in community care. The UK Medical Research Council's (MRC) framework for the development and evaluation of complex interventions [43,44] has guided the building of a rationale for adopting eHealth interventions in community care. The MRC's framework is recommended for the development of interventions containing several interacting components. The study reported in this paper pertains to the first step in the framework, which is Development. It encompasses identifying a relevant existing evidence base, ideally by carrying out a systematic review [44]. However, components of an intervention can also be identified through focus group interviews with the patients or health care personnel [45].

Aim and Research Questions

The aim of this study was to identify and explore the perspectives of the managers and health care personnel in community care about the use of eHealth interventions to prevent hospitalizations for home-dwelling older persons receiving community care.

There were 2 research questions that guided the study:

1. Which eHealth interventions do health care personnel identify as appropriate to apply to prevent avoidable hospitalizations of home-dwelling older persons receiving community care?
2. From the health care personnel's perspective, for which patients could hospitalizations potentially be prevented?

Methods

Context

The study was carried out in an urban municipality in Western Norway. Community care in this municipality is organized into 4 geographically based units and comprises 1600 older persons. This study involved 2 of these units with 800 older persons receiving community care. The municipality was in the process of integrating eHealth solutions in community care during the next few years.

This study was undertaken as a work package (WP) in a larger project, *Development and Implementation of eHealth in Municipalities*. The WP reported in this paper aimed at (1) identifying relevant patient groups who could potentially take advantage of eHealth in community care, (2) identifying the health care personnel's and managers' perspective of and readiness to use eHealth in community care, and (3) based on findings in (1) and (2), suggesting an eHealth intervention for the case municipality.

Design

The study design was a single embedded case study with a qualitative approach, comprising (1) individual interviews and (2) focus group interviews. A case study approach is particularly useful to employ when there is a need to obtain an in-depth appreciation of how eHealth could be used in community care to prevent hospitalizations for home-dwelling older persons, in its natural real-life context [46,47]. The case is defined as community care in a Norwegian municipality.

Recruitment and Data Collection

A total of 5 individual interviews and 2 focus group interviews (n=12) were undertaken to provide the health care personnel's and managers' perspective regarding the use of eHealth interventions, which potentially could prevent hospitalizations for home-dwelling older persons receiving community care. We conducted 5 individual semistructured interviews with senior managers in the municipality, applying a semistructured interview guide that focussed on the potential use of eHealth in community care. Using purposeful sampling [48], we sought informants who were most able to inform us on the research question. Senior managers were selected because they held major roles in the municipality's work with eHealth in community care and were in the best position to validate and provide relevant information for the study. Administrative personnel in the municipality who otherwise were not involved in this study recruited informants; they recommended potential informants who could best explicate the aspects of interest. MTG then asked potential informants face-to-face about participation and all accepted. There was no relationship between the informants and interviewer before study commencement. The interviews were conducted by the same person (MTG) for consistency and took place at the respective informants' office, with only the informant and interviewer present. The interviews lasted approximately 60 min and were audiotaped and transcribed verbatim.

We used focus group interviews to explore the health care personnel's perceptions of uptake and use of eHealth interventions that could potentially prevent hospitalizations. The focus group method is a useful data collection technique when the aim of the research is to explore attitudes, experiences, beliefs, and concerns [49]. In total, 2 focus group interviews (6+6 informants, n=12) were undertaken in 2014 by the author (MTG) as a moderator to ensure rich and relevant data [49]. A co-moderator made notes on observations and impressions during the interviews. Both interviews lasted approximately 90 min and were audiotaped and transcribed verbatim. A thematic interview guide was developed for the purpose of exploring aspects related to the uptake and use of eHealth interventions, including thoughts concerning which technological solutions they would have liked to have in their day-to-day practice and implementation and implications of eHealth in community care. To reduce the risk of any predetermined responses, participants did not see the interview guide before the interviews, thus also increasing the chance of open focus group discussions. To take advantage of homogeneity, shared experiences, and existing group dynamics, each focus group comprised health care personnel in direct patient care or nurse managers in community care. Administrative personnel in the municipality, who otherwise were not involved in this study, recruited informants. The studied municipality was in the process of integrating various technological applications in community care, consequently we were not in a position to seek participants who had operational experience of eHealth in their daily activities. To ensure appropriately experienced health care professionals working *on the ground*, we identified a maximum variation

sample; 12 health care professionals were invited and all agreed. Of them, 11 women and 1 man in the age range between 30 and 55 years, who had worked in community care for more than 5 years, participated in the interviews.

Data Analysis

Qualitative data were analyzed by way of systematic text condensation [50], as it is well-suited to analyze the multifaceted phenomena of eHealth. This approach involves the following steps in the analysis process: (1) establishing an overall impression of the data material and identifying preliminary themes, (2) identifying and sorting units of meaning into code groups, (3) condensing the contents of each of the code groups into subgroups, and (4) summarizing and recontextualizing the contents of each code group to generalize descriptions and concepts, in this case, related to the uptake and the use of eHealth in community care. Malterud argues that the data analysis will benefit from being conducted by more than one researcher [50], thus all authors read all interview transcripts to get an overall impression of the full data material, (step 1 of the systematic text condensation process). This step of the analysis requires the researcher to read, with an open mind from a bird's-eye perspective, all pages with transcripts and then ask which preliminary themes (usually 4-8 themes) can be identified in the material. We identified 4 preliminary themes: factors related to implementation, ethical aspects, training, and potential use.

This paper reports on findings related to the theme *potential use* (an analysis of contextual factors related to implementation has been published elsewhere [51]). The first author (MTG) undertook all the subsequent data analysis pertaining to the *potential use* theme (steps 2-4 of the systematic text condensation process) with input from the coauthors. The analytical process is demonstrated in Table 1.

Table 1. Analytical process.

Meaning units (selected)	Subgroups	Categories
Dehydrated; they are admitted for a short time, have some IV and then sent home.	Short stay; Simple intervention	Identification of potential patient groups
COPD patients are left to themselves when they are discharged, and then the anxiety comes...	Discharged without support	Identification of potential patient groups
A lot of UTIs...many men who are catheterized for 1,5-2 litres. If we had a bladder scanner, we could have solved it ourselves...instead of going to the A & E.	Potentially preventable hospitalization	Identification of potential patient groups
We don't have a bladder scanner, consequently we have to catheterize more often to be on the safe side, but then there is an infection and another hospitalization because of the infection	Use of technology to potentially prevent hospitalization	Identification of potential technological tool
A swollen leg, or whatever...there is much that could have been done if you could provide a picture or a video.	Video or photo as a tool for providing info about clinical condition	Identification of potential technological tool
... then we postpone, and eventually they are in such a bad shape that we have to call A & E.	—[a]	Identification of potential technological tool

[a]Not applicable.

Ethics

This project has been approved by the Norwegian Data Protection Official (approval ref# 21/2013). Informants have provided a written consent with information that they could redraw from the study at any point and without reason. Qualitative data from the interviews were transcribed verbatim and anonymized by exchanging informants' names with a number. We recorded informants' gender and years of work experience. All data were collected and stored in accordance with data protection regulations; stored electronically on computers, which were access-controlled via passwords. Hard copies of transcripts were securely stored in locked filing cabinets in offices that were accessible only to research staff. Data will be deleted at the end of the study.

Results

The data analysis of focus group interviews and individual interviews resulted in 2 categories: potential technological applications and potential patient groups. These 2 categories answer the research questions which eHealth interventions that are considered as appropriate to prevent hospitalizations, and a health care personnel's perspective on which patient groups hospitalizations potentially can be prevented. Content from step (4) in the analysis (recontextualization) is presented as analytical text with category heading, respectively, and assembled with quotes that are representative of the category.

Potential Technological Applications

Discussions in the focus groups generated several suggestions and wishes related to technical applications they could make use of in their day-to-day practice. Findings from the individual interviews identified several technology applications that could be useful in community care, but 1 manager expressed an important aspect:

> *It is very important to differentiate between the various types of technological applications; what can be useful in the day-to-day practice, for both the patients and the health care personnel in community care, in order to prevent hospitalizations and*

> *out-patient visits.* [Head of health and social welfare department]

The findings pertaining to this category demonstrates that the health care personnel warranted tools and measures to enhance and document their clinical observations in contact with patients. By doing so, they saw the potential of saving a trip to the outpatient emergency clinic for the patient and they could provide better quality of the home care.

In Norway, general practitioners (GPs) and doctors at the outpatient emergency clinic are obliged by law [52] to offer home visits to patients who are not fit to meet for a consultation at the doctor's clinic or when it is deemed necessary to provide sufficient treatment and care. The focus group participants described a practice where home visits were seldom undertaken, because either way the patient had to go to the clinic to take the necessary tests. If, however, home care personnel could have done the tests, they would have saved the patient for a potentially strenuous transportation to the doctor's clinic and at the same time reported much more precise clinical data. They had several suggestions in this matter; the possibility of drawing blood for a C-reactive Protein test (CRP) was suggested on the grounds that this was the first thing the doctor asked for when they made contact for an assessment of a patient. As they did not have the equipment to do this procedure, the patients had to book an appointment either with their GP or at the outpatient emergency clinic. According to the informants' experience, this often also involved the use of an ambulance for transportation. One situation they described was when they would contact the outpatient emergency clinic (in night-time and/or weekends) and they could only provide a diffuse description of the patient's condition, as they did not have access to measures that could help them be more precise in the description:

> *I'm calling the outpatient emergency clinic and report a patient who's had a general decline throughout the week, and the personnel there say that we have to take a blood sample (C-reactive Protein=CRP) and oxygen saturation... we can't perform this and consequently they are picked up by an ambulance.....We should have had the possibility to do these measures...* [Several nurses, focus group interview 2]

They also discussed the possibility of applying a video link to a doctor. This application could support their observations as well as provide a possibility for the doctor to assess a patient's condition without being face-to-face. They suggested using video link as a tool for the doctor to observe symptoms related to respiration and swollen legs/peripheral edema.

Furthermore, the participants in the focus group interviews wished to be equipped in a manner that made them more self-sufficient in providing high quality care and suggested, for example, the use of a bladder scanner as a tool, in relation to a problem with reoccurring urinary tract infections (UTIs):

> *If we had a bladder scanner, we could have solved it ourselves...instead of going to the outpatient emergency clinic.* [Nurse, focus group interview 1]

This was discussed in the context of patients who had been scanned and catheterized several times per day during their hospital stay (because of UTIs), whereas when they were discharged from the hospital to their home, the home care personnel had no tools to help them observe the phenomenon of residual urine. This is a crucial observation for the prevention of UTIs [53].

Informants in both the focus group interviews and individual interviews suggested the use of a tablet in the day-to-day practice in community care. A tablet installed with the quality/record system used in community care would enable the health care personnel to enhance and document their clinical observations in contact with patients. To date of the data collection, the personnel documented the clinical assessments on paper, which they would plot once they came to the home care base (office) where they had access to a computer and the patients' record. The informants expressed a clear potential to work safer, in terms of clinical measures being transmitted directly in the patients' record, as contrary to first record the measures on paper, bring it to the home care base, and then manually plot them in the record.

In the focus group interviews, the use of a tablet was also discussed as a means to be more prepared when there was an emergency callout. Emergency callouts were a daily activity, as most patients had a safety alarm that would alert the health care personnel in community care if they activated it (ie, pushed a pendant alarm). A typical situation would be if a patient had fallen, but it could also be that they were tired of waiting for their medication or wanted help to get to the toilet. However, the health care personnel would only receive an alarm signal, and the first step in the response was to receive a phone call from an emergency dispatcher who provided information about the patient's name, address, and phone number. If the alarm concerned a patient who they were not familiar with, they had no possibility to check the patient's record for relevant information:

> *One is always out driving, on the way from one patient to another, and then you have to stop the car, receive information about which patient – their name, address and phone number, by phone and write it down. It would be much easier to receive a text message with this information, and then log on to the patient's record on a tablet. I would like our quality system to be an app installed on a tablet!* [Nurse, focus group interview 2]

Informants both in the individual interviews and in the focus group interviews discussed technological applications related to a safe home environment and the potential for the patients to increase the degree of self-management using automated devices (smart house technology), alarms, and reminders. More concretely, they suggested that the safety alarm could be a hub for various types of applications, such as reminders for when to take their medication and when it was time to eat and movement-based light sensors located near the floor. The latter was suggested as a means to prevent patients from falling when they had to go to the toilet during the night.

Potential Patient Groups

Findings pertaining to this category represent a direct response to the question about patients for whom hospitalizations could be prevented. The findings stemming from the individual interviews bear a notion of managers being motivated by national policy regarding the resource utilization. The managers did not talk about specific patient groups but had a more general approach to preventing hospitalizations for home-dwelling older persons, which they described to be an appropriate task for the municipality/community care to undertake.

> *We have to look at possibilities for how to follow up on home-dwelling patients—they should not be admitted to hospital! We should be able to draw blood in their home and do measurements in their home...* [Assistant director]

The informants in the focus group interviews started off by discussing various clinical conditions, and patients that they viewed did not necessarily need the competence provided in specialized health care services that a hospital represents. If a patient was to be hospitalized because of dehydration, they considered the *treatment* or intervention initiated at the hospital to be rather short and simple, implying that this sort of intervention did not require specialized health care:

> *Dehydrated patients; they are admitted for a short time, have some intravenous fluid (IV) and then sent home.* [Nurse, focus group interview 2]

Another group of patients who they discussed about was those who have Chronic Obstructive Pulmonary Disease (COPD). In their experience, these patients had frequent readmissions to hospital, not necessarily because of the clinical condition itself, but because of the anxiety that often follows having respiratory problems:

> *COPD patients are left to themselves when they are discharged, and then the anxiety comes...the use of a telemonitoring device for promptly measures is neat.* [Nurse, focus group interview 1]

The informants agreed that a clinical condition described as potentially preventable was UTI. Especially, male patients were characterized as vulnerable in this context, as the personnel had to perform what they described as excessive catheterizations for men who had problems with residual urine in the bladder:

> *A lot of UTIs...many men who are catheterized for 1,5-2 liters. We have to catheterize more often to be on the safe side, but then there is an infection and then they are hospitalized again due to this...* [Nurse, focus group interview 1]

As demonstrated above, the informants in this study are quite clear about which technological applications they consider potentially useful in their practice. They also discussed the patient groups for whom hospitalizations potentially could be prevented. There was partly a connection between the suggested technological applications and the identified patient groups. The results are discussed against the relevant literature, providing suggestions for future intervention research.

Discussion

From a health care personnel's perspective, the main incentive to adopt eHealth in community care was the practical use in daily care. The various technological applications as well as different patient groups were identified, where the use of technological applications potentially could provide a more precise clinical assessment of home-dwelling older persons receiving care services.

Findings from this study revealed that the health care personnel in community care were vigilant in observing clinical decline but lacked tools to measure this decline. More specifically, they actually warranted the use of technological applications in their work, implying that they regard the use of eHealth as integral to their nursing practice in community care. This perspective is in contrast to what May et al found in their study from 2011 [2], where they identified problems in terms of health care professionals in community care to be indifferent and sometimes even hostile to the implementation of telecare systems. In addition, a more recent study by Greenhalgh et al [4] found that some clinicians would adopt readily to the use of video outpatient consultations, whereas others needed incentives and support.

However, May et al [2] also found that some health care professionals adopted the telecare service regardless, given that they perceived it as effective. On the basis of interviews with potential users of eHealth solutions (ie, health care personnel), our findings suggest that such applications have the potential to enable the nurses in community care to provide a more accurate description of the problem(s) when contacting a doctor. This implies that there is a potential to increase the quality of community care through the use of warranted technological applications. Moreover, the likelihood of successful adoption is increased as the interventions are considered to be useful and fit for purpose by the actual users [2,31,32]. Thus, the approach applied in our study provides great value in terms of developing appropriate interventions to prevent hospitalizations for home-dwelling older persons receiving community care, as it identifies issues and needs in practice [27,32,54].

The latter aspect is extremely important as it is increasingly recognized that the health care personnel's acceptance of the technological application itself remains a key challenge in adopting an intervention [28-30]. Furthermore, the informants in our study suggested the use of a video link to facilitate remote consultations with a doctor to deal with some of the nonurgent inquiries and potentially reduce the use of specialized health care services. Greenhalgh et al [4] found that video outpatient consultations appeared safe, effective, and convenient to use when the clinicians judged the patients to be clinically appropriate, but such situations were merely a fraction of the overall clinic workload. Although the use of a video link is perhaps not efficient in terms of reducing the workload, the informants in our study expressed an interest in saving patients for a potentially strenuous transportation to the doctor's clinic. This is an important care aspect, even though it cannot compromise the appropriate health care interventions to be undertaken. The finding must be seen in concordance with the

previous aspect; they saw the potential of providing improved quality of care, both in terms of making precise clinical assessments and caring for a patient´s resources. This holistic care practice is an expression of nurses who are well-grounded in general nursing knowledge, theory, practice competencies, and clinical experience and furthermore possess attributes of intuition and creativity to enhance a holistic care by the use of eHealth [42,55].

Previous research has identified that emergency hospitalizations of older persons often occur when the patient has reached a point of crisis, such as an exacerbation of a chronic condition, change in social setting, or a cascade of symptoms because of multimorbidity and frailty [14-17]. The use of technological applications as suggested by the health care personnel in our study could potentially prevent a severe state of illness that requires hospitalization, by discovering and addressing the patients' symptoms at an early stage. This is important with regard to both quality of care and resource utilization [13,19].

Limitations

This case study does not formulate a solution for how an eHealth intervention should be developed, but the insights from the study could inform a future intervention in comparable settings. One premise in this paper is to acknowledge that people and technologies are linked in a dynamic health care system made up of multiple interacting stakeholders. We have not focused on the patients or other stakeholders (eg, technology suppliers) as intended users of a technological solution. This needs to be explored for building an even more solid rationale for applying a technological application in community care. An intervention should be informed by all stakeholders—individual users, service providers, and technology suppliers— to ensure a person-centered, holistic, and ethical approach. Such coproduction should be addressed in future research.

The findings from this case study pertain to a particular community care and context prevailing in the included Norwegian municipality. Other municipalities, countries, and settings may illustrate different opportunities and challenges, which should be explored. It could be argued that our sample of informants including 17 community care managers and health care personnel should have been larger. However, based on the study's rather narrow aim and the use of theory to extend the sources of knowledge beyond the empirical interview data, the sample offered sufficient information power, as described by Malterud et al [56]. The sample of 12 health care personnel had daily patient contact and represented future users of eHealth solutions. Hence, their perspectives may be transferable to other similar contextual settings as described in this study.

Conclusions

Through this study, we have generated empirical knowledge about which eHealth interventions could potentially prevent hospitalizations for home-dwelling older persons receiving community care. By identifying issues and needs in practice we have identified factors paramount for acceptance and adoption of an intervention [27,54]. We have shown that the health care personnel in community care warrant various technological applications that have the potential to improve quality of care and resource utilization in the studied municipality.

Previous research has pointed to a poorly founded rationale for the use of an eHealth intervention as a reason for slow and fragmented uptake and use of eHealth in community care [27,41]. The findings in this study can specifically inform future interventions aiming to prevent hospitalizations for home-dwelling older persons in community care, as the identified potential applications are considered useful and fit for purpose. Furthermore, by providing a description of the development phase of a future intervention as described in the MRC's framework [44], it adds significantly to the general body of knowledge regarding developing eHealth interventions in community care.

Acknowledgments

The study was funded by the Centre for Age-Related Medicine, Stavanger University Hospital; the Western Regional Research Fund through Stavanger municipality; and the University of Stavanger. The authors would like to thank all the informants for their invaluable contribution to the study. The authors would also like to thank Henrik Hovland, who assisted in conducting focus group interviews.

Authors' Contributions

MTG planned the study design, was responsible for the development of data collection tools, contributed to data analysis, and drafted this manuscript. SW contributed to the study design, contributed to the development of data collection tools, data analysis, and contributed to drafting the manuscript. IT contributed to data analysis and drafting the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. United Nations Population Fund (UNFPA). Ageing in the Twenty-First Century: A Celebration and a Challenge. New York: United Nations Population Fund; Helpage International; 2012.

2. May C, Finch T, Cornford J, Exley C, Gately C, Kirk S, et al. Integrating telecare for chronic disease management in the community: what needs to be done? BMC Health Serv Res 2011 May 27;11:131 [FREE Full text] [doi: 10.1186/1472-6963-11-131] [Medline: 21619596]
3. Lewin D, Adshead S, Glennon B, Williamson B, Moore T, Damodaran L, et al. Michael McGetrick PhD. London: Plum Consulting; 2010 Mar. Assisted Living Technologies for Older and Disabled People in 2030: A Final Report to Ofcom URL: http://www.drmichaelmcgetrick.com/ALT_Report.pdf [accessed 2019-11-21]
4. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
5. Peek ST, Wouters EJ, Luijkx KG, Vrijhoef HJ. What it takes to successfully implement technology for aging in place: focus groups with stakeholders. J Med Internet Res 2016 May 3;18(5):e98 [FREE Full text] [doi: 10.2196/jmir.5253] [Medline: 27143097]
6. European Commission. Digital Agenda for Europe. Rebooting Europe's Economy. Brussels: Directorate-General for Communication Citizens Information; 2014.
7. Kang HG, Mahoney DF, Hoenig H, Hirth VA, Bonato P, Hajjar I, Center for Integration of Medicine and Innovative Technology Working Group on Advanced Approaches to Physiologic Monitoring for the Aged. In situ monitoring of health in older adults: technologies and issues. J Am Geriatr Soc 2010 Aug;58(8):1579-1586. [doi: 10.1111/j.1532-5415.2010.02959.x] [Medline: 20646105]
8. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. Br Med J 2012 Jun 21;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
9. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
10. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ Open 2013 Aug 19;3(8):e003482 [FREE Full text] [doi: 10.1136/bmjopen-2013-003482] [Medline: 23959760]
11. Ellis G, Whitehead MA, Robinson D, O'Neill D, Langhorne P. Comprehensive geriatric assessment for older adults admitted to hospital: meta-analysis of randomised controlled trials. Br Med J 2011 Oct 27;343:d6553 [FREE Full text] [doi: 10.1136/bmj.d6553] [Medline: 22034146]
12. Strømgaard S, Rasmussen SW, Schmidt TA. Brief hospitalizations of elderly patients: a retrospective, observational study. Scand J Trauma Resusc Emerg Med 2014 Mar 7;22:17 [FREE Full text] [doi: 10.1186/1757-7241-22-17] [Medline: 24606987]
13. Klausen HH, Petersen J, Bandholm T, Juul-Larsen HG, Tavenier J, Eugen-Olsen J, et al. Association between routine laboratory tests and long-term mortality among acutely admitted older medical patients: a cohort study. BMC Geriatr 2017 Mar 1;17(1):62 [FREE Full text] [doi: 10.1186/s12877-017-0434-3] [Medline: 28249621]
14. Crane SJ, Tung EE, Hanson GJ, Cha S, Chaudhry R, Takahashi PY. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. BMC Health Serv Res 2010 Dec 13;10:338 [FREE Full text] [doi: 10.1186/1472-6963-10-338] [Medline: 21144042]
15. Gamper G, Wiedermann W, Barisonzo R, Stockner I, Wiedermann CJ. Inappropriate hospital admission: interaction between patient age and co-morbidity. Intern Emerg Med 2011 Aug;6(4):361-367. [doi: 10.1007/s11739-011-0629-0] [Medline: 21655929]
16. Philp I, Mills KA, Thanvi B, Ghosh K, Long JF. Reducing hospital bed use by frail older people: results from a systematic review of the literature. Int J Integr Care 2013;13:e048 [FREE Full text] [doi: 10.5334/ijic.1148] [Medline: 24363636]
17. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet 2013 Mar 2;381(9868):752-762 [FREE Full text] [doi: 10.1016/S0140-6736(12)62167-9] [Medline: 23395245]
18. Sánchez-García S, Juárez-Cedillo T, Mould-Quevedo J, García-González JJ, Contreras-Hernández I, Espinel-Bermudez MC, et al. The hospital appropriateness evaluation protocol in elderly patients: a technique to evaluate admission and hospital stay. Scand J Caring Sci 2008 Jun;22(2):306-313. [doi: 10.1111/j.1471-6712.2007.00528.x] [Medline: 18489701]
19. D'Souza S, Guptha S. Preventing admission of older people to hospital. Br Med J 2013 May 20;346:f3186. [doi: 10.1136/bmj.f3186] [Medline: 23690461]
20. Rechel B, Grundy E, Robine JM, Cylus J, Mackenbach JP, Knai C, et al. Ageing in the European union. Lancet 2013 Apr 13;381(9874):1312-1322. [doi: 10.1016/S0140-6736(12)62087-X] [Medline: 23541057]
21. Huseby BM. [Samdata Specialist Health Service 2013]. Oslo: The Norwegian Directorate of Health; 2014.
22. Roland M, Abel G. Reducing emergency admissions: are we on the right track? Br Med J 2012 Sep 18;345:e6017. [doi: 10.1136/bmj.e6017] [Medline: 22990102]
23. Deraas TS, Berntsen GR, Jones AP, Førde OH, Sund ER. Associations between primary healthcare and unplanned medical admissions in Norway: a multilevel analysis of the entire elderly population. BMJ Open 2014 Apr 12;4(4):e004293 [FREE Full text] [doi: 10.1136/bmjopen-2013-004293] [Medline: 24727427]

24. Bloom DE, Chatterji S, Kowal P, Lloyd-Sherlock P, McKee M, Rechel B, et al. Macroeconomic implications of population ageing and selected policy responses. Lancet 2015 Feb 14;385(9968):649-657 [FREE Full text] [doi: 10.1016/S0140-6736(14)61464-1] [Medline: 25468167]
25. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. Lancet 2009 Oct 3;374(9696):1196-1208 [FREE Full text] [doi: 10.1016/S0140-6736(09)61460-4] [Medline: 19801098]
26. Forsiden. Report No. 47 (2008-2009) to the Storting: The Cooperation Reform. Proper treatment - at the right place and right time URL: https://www.regjeringen.no/en/dokumenter/report.no.-47-to-the-storting-2008-2009/id567201/ [accessed 2013-08-01]
27. Taylor J, Coates E, Brewster L, Mountain G, Wessels B, Hawley MS. Examining the use of telehealth in community nursing: identifying the factors affecting frontline staff acceptance and telehealth adoption. J Adv Nurs 2015 Feb;71(2):326-337. [doi: 10.1111/jan.12480] [Medline: 25069605]
28. Hendy J, Chrysanthaki T, Barlow J, Knapp M, Rogers A, Sanders C, et al. An organisational analysis of the implementation of telecare and telehealth: the whole systems demonstrator. BMC Health Serv Res 2012 Nov 15;12:403 [FREE Full text] [doi: 10.1186/1472-6963-12-403] [Medline: 23153014]
29. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. J Adv Nurs 2014 Jan;70(1):21-33. [doi: 10.1111/jan.12196] [Medline: 23786584]
30. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
31. Gjestsen MT, Wiig S, Testad I. What does it take? Healthcare professional's perspective on incentives and obstacles related to implementing ICTs in home-based elderly care. BMC Health Serv Res 2014;14(S2):45. [doi: 10.1186/1472-6963-14-s2-p45]
32. Joseph V, West RM, Shickle D, Keen J, Clamp S. Key challenges in the development and implementation of telehealth projects. J Telemed Telecare 2011;17(2):71-77. [doi: 10.1258/jtt.2010.100315] [Medline: 21097563]
33. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015;13(1):91. [doi: 10.1186/s12916-015-0279-6]
34. Koch S, Hägglund M. Health informatics and the delivery of care to older people. Maturitas 2009 Jul 20;63(3):195-199. [doi: 10.1016/j.maturitas.2009.03.023] [Medline: 19487092]
35. Bowes A, McColgan G. Telecare for older people: Promoting independence, participation, and identity. Res Aging 2013;35(1):32-49. [doi: 10.1177/0164027511427546]
36. Cohn S, Clinch M, Bunn C, Stronge P. Entangled complexity: why complex interventions are just not complicated enough. J Health Serv Res Policy 2013 Jan;18(1):40-43. [doi: 10.1258/jhsrp.2012.012036] [Medline: 23393041]
37. 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]
38. Aarts J, Gorman P. IT in health care: sociotechnical approaches 'To Err is System'. Int J Med Inform 2007;76(1):S1-S3. [doi: 10.1016/s1386-5056(07)00078-0]
39. Li J. A sociotechnical approach to evaluating the impact of ICT on clinical care environments. Open Med Inform J 2010 Sep 15;4:202-205 [FREE Full text] [doi: 10.2174/1874431101004010202] [Medline: 21594005]
40. Berg M, Aarts J, van der Lei J. ICT in health care: sociotechnical approaches. Methods Inf Med 2003;42(4):297-301. [doi: 10.1267/meth03040297] [Medline: 14534625]
41. Greenhalgh T, Shaw S, Wherton J, Hughes G, Lynch J, A'Court C, et al. SCALS: a fourth-generation study of assisted living technologies in their organisational, social, political and policy context. BMJ Open 2016 Feb 15;6(2):e010208 [FREE Full text] [doi: 10.1136/bmjopen-2015-010208] [Medline: 26880671]
42. Potter PJ, Frisch N. Holistic assessment and care: presence in the process. Nurs Clin North Am 2007 Jun;42(2):213-28, vi. [doi: 10.1016/j.cnur.2007.03.005] [Medline: 17544679]
43. 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]
44. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
45. Richards DA, Hallberg IR. Complex Interventions in Health: An overview of research methods. London: Routledge; 2015.
46. Yin RK. Case Study Research : Design and Methods. Fifth Edition. London: Sage; 2014.
47. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
48. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: Sage Publications; 2014.
49. Morgan DL. Focus Groups as Qualitative Research. Second Edition. Thousand Oaks, CA: Sage Publications; 1997.

50. Malterud K. Systematic text condensation: a strategy for qualitative analysis. Scand J Public Health 2012 Dec;40(8):795-805. [doi: 10.1177/1403494812465030] [Medline: 23221918]
51. Gjestsen MT, Wiig S, Testad I. What are the key contextual factors when preparing for successful implementation of assistive living technology in primary elderly care? A case study from Norway. BMJ Open 2017 Sep 7;7(9):e015455 [FREE Full text] [doi: 10.1136/bmjopen-2016-015455] [Medline: 28882908]
52. Norwegian Ministry of Health and Care Services Services. Law data: The front page. Regulations on Regular Medical Scheme in the Municipalities URL: https://lovdata.no/dokument/SF/forskrift/2012-08-29-842 [accessed 2018-06-20]
53. Truzzi JC, Almeida FM, Nunes EC, Sadi MV. Residual urinary volume and urinary tract infection--when are they linked? J Urol 2008 Jul;180(1):182-185. [doi: 10.1016/j.juro.2008.03.044] [Medline: 18499191]
54. Broderick A, Lindeman D. SAMHSA-HRSA. 2013. Scaling telehealth programs: Lessons from early adopters. Case studies in telehealth adoption URL: https://www.integration.samhsa.gov/operations-administration/Telehealth_Commonwealth_fund_Report_1654_Broderick_telehealth_adoption_synthesis.pdf [accessed 2018-06-13]
55. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
56. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]

Abbreviations

COPD: chronic obstructive pulmonary disease
CRP: C-reactive protein
eHealth: electronic health
GP: general practitioner
MRC: Medical Research Council
UTI: urinary tract infection
WP: work package

Edited by G Eysenbach; submitted 12.11.18; peer-reviewed by K Mishina, K Anderson; comments to author 27.12.18; revised version received 30.04.19; accepted 03.06.19; published 02.01.20.

Please cite as:
Gjestsen MT, Wiig S, Testad I
Health Care Personnel's Perspective on Potential Electronic Health Interventions to Prevent Hospitalizations for Older Persons Receiving Community Care: Qualitative Study
J Med Internet Res 2020;22(1):e12797
URL: https://www.jmir.org/2020/1/e12797
doi:10.2196/12797
PMID:31895045

Review

Effectiveness of Upper Limb Wearable Technology for Improving Activity and Participation in Adult Stroke Survivors: Systematic Review

Jack Parker[1], BSc, PhD; Lauren Powell[2], MSc, BSc (Hons); Susan Mawson[2], BSc (Hons), MCSP, PhD

[1]University of Derby, Derby, United Kingdom

[2]University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Lauren Powell, MSc, BSc (Hons)
University of Sheffield
Western Bank
Sheffield, S10 2TN
United Kingdom
Phone: 44 01142228179
Email: l.a.powell@sheffield.ac.uk

Abstract

Background: With advances in technology, the adoption of wearable devices has become a viable adjunct in poststroke rehabilitation. Upper limb (UL) impairment affects up to 77% of stroke survivors impacting on their ability to carry out everyday activities. However, despite an increase in research exploring these devices for UL rehabilitation, little is known of their effectiveness.

Objective: This review aimed to assess the effectiveness of UL wearable technology for improving activity and participation in adult stroke survivors.

Methods: Randomized controlled trials (RCTs) and randomized comparable trials of UL wearable technology for poststroke rehabilitation were included. Primary outcome measures were validated measures of activity and participation as defined by the International Classification of Functioning, Disability, and Health. Databases searched were MEDLINE, Web of Science (Core collection), CINAHL, and the Cochrane Library. The Cochrane Risk of Bias Tool was used to assess the methodological quality of the RCTs and the Downs and Black Instrument for the quality of non RCTs.

Results: In the review, we included 11 studies with collectively 354 participants at baseline and 323 participants at final follow-up including control groups and participants poststroke. Participants' stroke type and severity varied. Only 1 study found significant between-group differences for systems functioning and activity ($P \leq .02$). The 11 included studies in this review had small sample sizes ranging from 5 to 99 participants at an average (mean) age of 57 years.

Conclusions: This review has highlighted a number of reasons for insignificant findings in this area including low sample sizes and the appropriateness of the methodology for complex interventions. However, technology has the potential to measure outcomes, provide feedback, and engage users outside of clinical sessions. This could provide a platform for motivating stroke survivors to carry out more rehabilitation in the absence of a therapist, which could maximize recovery.

Trial Registration: PROSPERO CRD42017057715; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=57715

(J Med Internet Res 2020;22(1):e15981) doi:10.2196/15981

KEYWORDS

wearable electronic devices; stroke; rehabilitation; upper extremity

Introduction

Background

Stroke is a leading cause of mortality and disability worldwide [1], and the economic costs of treatment and poststroke care result in a mean cost to the economy of £46,039 a year per patient for the first 5 years after admission in England, Wales, and Northern Ireland alone [2]. Following a stroke, survivors are left with multiple impairments, and as a result only 5% to 20% will regain full function of the upper limb (UL) with up

to 66% still being impaired in the chronic phase [3]. This often results in functional limitations in activities of daily living and decreased quality of life [4].

Over recent years, there has been a contextual shift in service delivery from hospital-based rehabilitation to the community. Although it has been recommended that rehabilitation should continue until maximum recovery has been achieved [5], owing to the increasing demand on services and financial constraints, service needs cannot be met; therefore, radical innovation and the adoption of a self-management paradigm are considered as a way of delivering independent home-based rehabilitation, thereby meeting the challenges faced in health care [6].

Evidence exists supporting the need for intensity and repetition of motor skills to promote neuroplasticity and motor relearning. With significant advances in information and communication technology (ICT) and more specifically in the rapid development and deployment of sensor technology for health care monitoring, a number of technological aids with a potential to measure and monitor poststroke activity have been explored for both the lower limb [7] and the UL [8]. However, many include the use of expensive, large, complex, ungainly equipment that is impractical to use in everyday contexts [9]. Therefore inexpensive, wearable, and commercially available sensors have become a more viable option for quantifying movements and activities during poststroke rehabilitation [10-12].

A number of recent systematic and nonsystematic reviews highlight the growing use of wearable devices to provide poststroke rehabilitation in both clinical and nonclinical settings for motion analysis and physical activity monitoring [12-17]. These include microelectromechanical systems containing accelerometers, gyroscopes, and magnetometers; fabric and body-worn sensor networks [18], pressure sensors [19-22], and physiological monitoring such as blood pressure and oxygen saturation [23,24]. Other wearable devices specifically designed and used for poststroke rehabilitation also include robotics [25], virtual reality [26], functional electrical stimulation (FES) [27,28], electromyographic biofeedback [29], and transcutaneous electrical nerve stimulation [30-32].

However, while these devices have the potential to reliably measure duration, frequency, intensity, and quality of activity and movement, all of which are key variables for poststroke recovery [33], no reviews have synthesized the evidence underpinning the use of these devices for independent poststroke UL rehabilitation. Therefore, the aim of this review will be to explore and examine how effective these medical devices are as interventions for improving the function of the UL in adult stroke survivors.

The International Classification of Functioning, Disability, and Health (ICF) [34] considers the interaction between pathology (body structure and function), impairment (signs and symptoms), activities (functionality), and participation (social integration), and it has now become the main conceptual framework for poststroke rehabilitation [5,35,36].

Objective

For this review, we focused on the activities and participation domain of the ICF as this would provide an indication of how the interventions have or have not led to functional gains in everyday life.

Methods

The review protocol was registered on PROSPERO (CRD42017057715). The review was undertaken in accordance with the general principles recommended in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [37].

Definitions

Wearable technologies can be subdivided into those operating independently and functioning as central connectors for other devices “and” or “or” information (eg, wrist-worn fitness tracker and smartphone) and those capturing specific actions or executing a measurement (eg, heart rate monitor worn around the chest) offloading to a primary wearable device for analysis [38]. We define a wearable device in the context of poststroke rehabilitation as “a wearable device that is worn externally on the body that is portable (the user is able to wear the device but is free to move around and not fixed to a station) and is able to use the device independently of a therapist.”

Search Methods

As per the Cochrane Handbook [39], the Population Intervention Comparison Outcome Study Design framework helped authors to define the inclusion and exclusion criteria and the search terms of this review. For this review, the population refers to poststroke adults, the intervention to technological interventions for UL rehabilitation in stroke survivors, and included studies included a comparison group and were not limited to randomized controlled trials (RCTs). The outcome focused on activity and participation measures of UL function poststroke. Search terms and databases were selected based on Cochrane literature and institutional information specialist advice.

The following databases were searched from the year 2000 to April 2019: MEDLINE, Web of Science (Core collection), CINAHL, Scopus, and the Cochrane Library. Medical Subject Headings (MeSH) keywords used were cerebrovascular disorders, hemorrhage, cerebral hemorrhage, sensory feedback, motor skills, physical therapy modalities, physical and rehabilitation medicine, exercise, exercise therapy, rehabilitation, exercise movement techniques, information technology, technology, self-help devices, telemedicine, upper extremity, arm, hand joints, shoulder joint, elbow joint, and wrist joint. Text terms used were stroke, UL, rehabilitation, and technology. These were combined with the following synonyms: CVA, cerebrovascular accident, poststroke, cerebrovascular, brain ischemia, brain vascular, upper extremity, arm, shoulder, hand, axilla, axilla, elbow, forearm, finger, wrist, physiotherapy, physical therapy, physiatric, exercise, biofeedback, sensory feedback, advise, train, therapy, treat, motor skills, motor relearn, re-educate, recovery, enhance, promote, support, function, activity, physical, information technology, IT, ICT, information and communications technology, assistive technology, telehealth, telecare, telerehabilitation, and wear. Boolean logic was used to combine terms using *AND* and *OR*. MeSH terms refer to specific terms that are recognized for

indexing journals and books in electronic databases. The free text terms and synonyms were words used in the search strategy that was looked for in titles and abstracts. The MEDLINE search strategy can be found in Multimedia Appendix 1. Electronic citations were downloaded into a reference manager. The inclusion and exclusion criteria for the search strategy are presented in Textboxes 1 and 2, respectively.

As technology is changing very quickly, authors deemed technology before the year 2000 to be particularly outdated. RCTs and randomized comparable trials were chosen as the appropriate study design for inclusion in this review as the review aims to assess the effectiveness of the included interventions. Non-RCT and nonrandomized comparable trial evidence is therefore outside the scope of this review. Comparators (control groups) could include treatment as usual and exercise therapies that do not include any other intervention or sham stimulation.

The primary outcome measures for this review are those that assess activity or participation as defined by the World Health Organization (WHO) ICF [40]. These measures include the following: the Box and Blocks Test (BBT) [41]; Action Research Arm Test (ARAT) [42]; Barthel Index (BI) [43]; Chedoke Arm and Hand Activity Inventory (CAHAI) [44-47]; Jebson-Taylor Hand Function Test (JTHFT) [48]; Wolf Motor Function Test (WMFT) [49]; Motor Activity Log (MAL) [50]; Motor Assessment Scale (MAS) [51]; Stroke Impact Scale (SIS) [52]; the Rivermead Motor Assessment (RMA) [53]; Upper Extremity Function Test (UEFT) [54]; and the short version of disabilities of arm, shoulder, and hand (QuickDASH) [55].

A total of 3 measures of *system functioning* WHO ICF, namely, the Fugl-Meyer Test [56], the Arm Motor Ability Test [57], and the pain Visual Analogue Scale [58], were not included in this review as the aim was to explore and examine how effective medical devices are used as interventions for improving function (activity and participation) of the UL in adult stroke survivors.

Textbox 1. Inclusion criteria.

- English-language articles
- Studies recruiting people over the age of 18 years
- Studies evaluating upper limb wearable technology
- Studies reporting randomized controlled trials or randomized comparable trials
- Studies measuring activity and or participation as classified by the World Health Organization International Classification of Functioning, Disability, and Health
- Intervention that could be used independently by the stroke survivor
- Wearable and portable technology that measures or monitors activity
- Research article

Textbox 2. Exclusion criteria.

- Non-English–language articles
- Studies recruiting people under the age of 18 years
- Studies not evaluating upper limb wearable technology
- Studies not reporting randomized controlled trials or randomized comparable trials
- Studies not measuring activity and or participation as classified by the World Health Organization International Classification of Functioning, Disability, and Health
- Intervention that could not be used independently by the stroke survivor
- Wearable and portable technology that does not measure or monitors activity
- Not a research article
- Studies where the intervention is not clearly defined (it was unclear to the authors that the study did or did not meet the inclusion/exclusion criteria)
- Study protocols
- Studies reporting a nonwearable, nonportable intervention

Quality Assessment of Included Studies

The methodological quality of the included RCTs was assessed using the Cochrane Risk of Bias (CRoB) assessment criteria [59]. This addresses specific fields including sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting. RCTs were classed as having an overall low risk of bias if they were rated as *low* for 3 of the key areas: allocation concealment [60], blinding of outcome assessment, and completeness of outcome data. They were judged as overall high risk of bias if any of these key areas

were judged as being an overall *high* risk. RCTs judged as being at an overall unclear risk of bias were so if any of the 3 areas above were judged as *unclear*.

For the included non-RCTs, the methodological quality was assessed using the Downs and Black Instrument [61]. This instrument provides a score for each study, and the maximum score is 37. It assesses the way studies report their findings, their external and internal validity as well as selection bias.

Data Extraction

The titles, abstracts, and/or papers were screened by the authors LP and JP to find studies that meet the review inclusion criteria. Final papers were decided between the authors JP and LP, and any disagreement was resolved through discussion. A standardized Excel form was used to extract data and study characteristics. This is where information such as data on the interventions and participant characteristics were recorded. The author LP carried out the data extraction and checked for accuracy by the author JP. Whenever applicable, missing data were requested from the authors of the study.

Outcome Measure Quality Assessment

When undertaking a systematic review, it is important to assess the quality of the outcome measures used in the included studies to ensure that the results are valid and reliable. To achieve this, 3 clear domains can be considered for each of the outcome measures used: (1) whether the psychometric properties of the scale have been assessed previously [62], (2) whether the clinimetric properties of the scale have been considered [63-67], specifically the Minimally Clinically Important Difference (MCID) [66], and (3) whether the statistical analysis of the data provided by the scale fulfills the requirements of measurement theory [68-70].

We identified all outcome measures (N=12) used in the 11 included studies and reviewed each individually to assess whether they fulfilled the first 2 domains outlined above. The outcome measures measuring activity included BBT [41]; ARAT [42]; BI [43]; CAHAI [44-47]; JTHFT [48]; WMFT [49]; MAL [50]; MAS [51]; RMA [53]; UEFT [54]; and QuickDASH [55]; and the outcome measures measuring participation included SIS [52].

This was determined by reviewing the literature of each of the outcome measures. How the outcome measure was used and how the data were scored and analyzed was then examined for each of the 11 included studies.

All 12 outcome measures were measures of activity (N=11) or participation (N=1) as classified by the WHO ICF [40].

Data Synthesis

A narrative review is presented on the included studies with supporting evidence tables for study characteristics and findings, risk of bias, and outcome measure assessment. A meta-analysis was not undertaken because of the variability of outcome measures used across the 12 included studies.

Appraisal of Evidence

Studies in this review include RCTs and randomized trials without a control group. The included studies were appraised using the levels of evidence [71]. This is to enhance the understanding of the best levels of evidence included in this review [72]. The highest level of evidence to this end is that of the RCT and is considered to be of *level 1* evidence. Randomized trials without a control group are considered to be of *level 2* evidence. This is important as the study design can affect the validity and reliability of results. For example, RCTs are often considered the *gold standard* of research evidence and the most reliable because of the measures taken to reduce the influences of confounds [73].

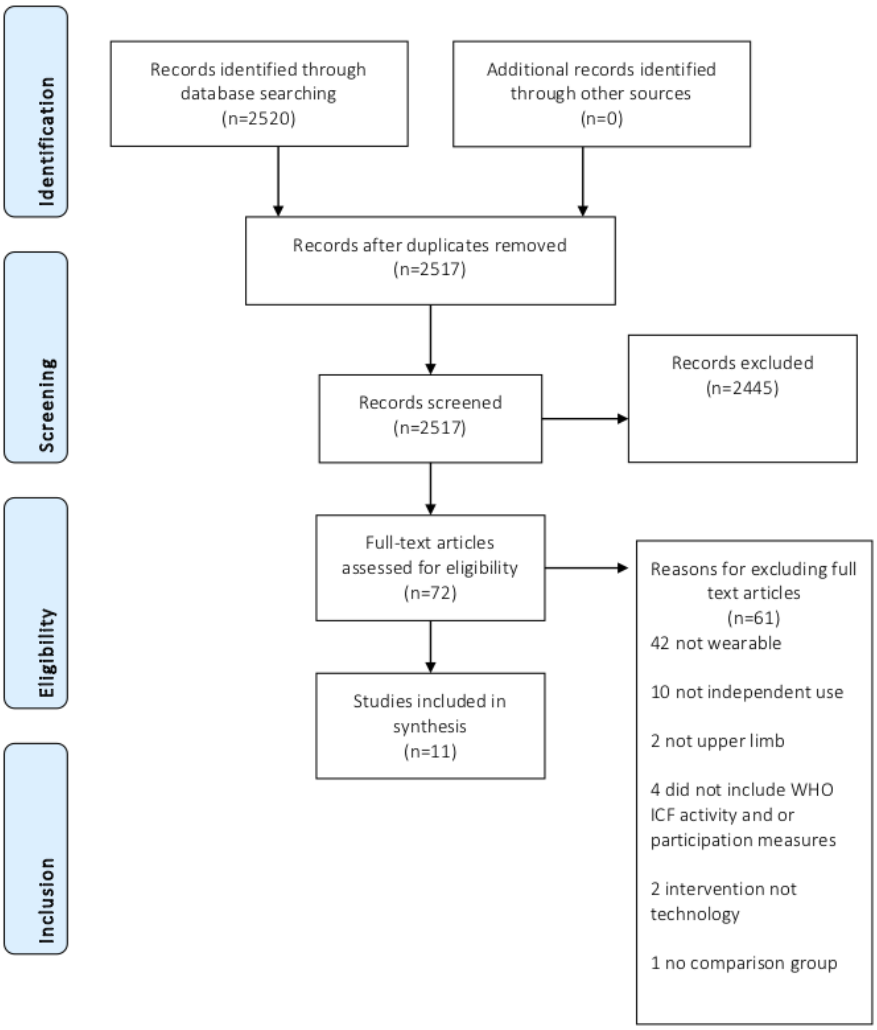
Results

Search Results

The electronic searches identified 2517 citations following deduplication. No additional citations were identified via handpicking methods. Following deduplication, 2517 records were screened and 2445 records were excluded through the title and abstract screening phases. At this stage, 72 full papers were obtained and 61 of these were excluded (reasons for exclusion can be found in Multimedia Appendix 2 [74-132]). Of these, 11 studies reported across 11 publications were included in this review (see Figure 1).

The review included 11 studies carried out in the United States (5), the Netherlands (2), Australia (1), Spain (1), Turkey (1), and Italy (1) with collectively 354 participants at baseline and 323 participants at final follow-up including control groups and participants from 17 days to 5-year poststroke. Of which, 7 studies were RCT level 1 and 4 were level 2 comparison trials and 6 of the 11 studies included acute stroke survivors. Participants' stroke type and severity varied from mild to severe. The interventions used FES (3), a hand device/glove (7), and arm worn garment (1). The duration of the intervention was from 3 to 12 weeks with varying intensity. Only 1 study found significant between-group differences for systems functioning and activity ($P \leq .02$). The 11 included studies in this review had small sample sizes ranging from 5 to 99 participants at an average (mean) age of 57 years old.

Figure 1. Article selection. WHO ICF: The World Health Organization International Classification of Functioning, Disability, and Health.



Quality Assessment

The CRoB quality assessment summary can be found in Table 1, the Downs and Black quality assessment for non-RCT designs in Table 2 and the outcome measure quality assessment in Multimedia Appendix 3. Full details of the CRoB assessment can be found in Multimedia Appendix 4 and the Downs and Black quality assessment in Multimedia Appendix 5. First, 2 of the 7 included RCTs were judged as having an overall high risk of bias [73]. Of these, 1 of these RCTs was judged to be at high risk of having incomplete outcome data [141], 1 reported outcome assessment was not blinded [142], and 3 did not report blinding of participants and personnel [142-144]. Then, 4 RCTs were judged as having an overall unclear risk of bias [144-147], and 1 of the included RCTs was considered to be at an overall low risk of bias [143].

Non-RCT evidence was assessed using the Downs and Black Instrument, as they were studies that did not involve control groups. Overall, the 4 studies assessed using this instrument received high scores for reporting domains (items 1-10) and internal validity bias (items 14-20). Overall, they obtained lower scores for the external validity (items 11-13), selection bias (items 21-26), and power (item 27). The maximum score that could be obtained from this instrument is 32. Of the 4 studies assessed using this method, the highest score obtained was 21 [148] and the lowest was 14 [149].

Table 1. Cochrane risk of bias quality assessment summary.

Author, year	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Overall judgment
Alon, 2008 [145]	Unclear	Unclear	Unclear	Unclear	Unclear	Low risk	Unclear
da Silva Cameira˜o, 2011 [141]	Unclear	Unclear	Unclear	Low risk	High risk	Low risk	High risk
Lannin, 2016 [143]	Low risk	Low risk	High risk	Low risk	Low risk	High risk	Low risk
Nijenhuis, 2017 [142]	Low risk	Unclear	High risk	High risk	Low risk	Low risk	High risk
Vilafane, 2018 [147]	Low risk	Unclear	Low risk	Low risk	Low risk	Low risk	Unclear
Wolf, 2015 [144]	Low risk	Unclear	High risk	Low risk	Low risk	Low risk	Unclear
Nakipoglu Yuzer, 2017 [146]	Low risk	Unclear	Unclear	Unclear	Low risk	Low risk	Unclear

Table 2. Downs and Black quality assessment summary.

	Author, year			
	Barry, 2012 [150]	Friedman, 2014 [148]	Knutson, 2016 [151]	Prange-Lasonder 2017 [149]
Reporting				
1[a]	1	1	1	1
2[b]	1	1	1	1
3[c]	1	1	1	1
4[d]	1	1	1	1
5[e]	2	2	2	2
6[f]	1	1	1	1
7[g]	1	1	1	0
8[h]	0	0	1	0
9[i]	1	0	1	0
10[j]	1	1	1	1
External validity				
11[k]	0	1	0	0
12[l]	0	1	0	0
13[m]	0 UTD[n]	0 UTD	0 UTD	0 UTD
Internal validity—bias				
14[o]	0	1	0	0
15[p]	1	1	1	0
16[q]	1	1	1	1
17[r]	0 N/A[s]	1	1	1
18[t]	1	1	1	1
19[u]	0 UTD	1	1	0 UTD
20[v]	1	1	1	1
Internal validity—confounding (section bias)				
21[w]	0 UTD	1	1	1
22[x]	0 UTD	0 UTD	0 UTD	0 UTD
23[y]	1	1	1	1
24[z]	0 UTD	1	0 UTD	0 UTD
25[aa]	0 UTD	0 UTD	0 UTD	0 UTD
26[ab]	1	0 UTD	1	0 UTD
Power				
27[ac]	0	0	0	0
Total	16	21	20	14

[a]1: Clarity of aims, objectives, and hypothesis.

[b]2: Clarity of main outcomes described.

[c]3: Clarity of participant characteristics.

[d]4: Clarity of intervention description.

[e]5: Clarity of distributions of principal confounders in each group.

[f]6: Are the main findings of the study clearly described?

[g]7: Are estimates of random variability in data for main outcomes clearly described?

[h]8: Have all adverse effects related to the intervention been reported?

[i]9: Have lost to follow-up participant characteristics been described?

[j]10: Have probability values for main outcomes been reported except from where $P<.001$?

[k]11: Were the participants asked to take part in the study representative of the entire population from which they were recruited?

[l]12: Were the participants prepared to take part in the study representative of the population from which they were recruited?

[m]13: Were the staff, places, and facilities where the participants were treated representative of the treatment that the majority of patients receive?

[n]UTD: unable to determine.

[o]14: Was there an attempt to blind participants?

[p]15: Was there an attempt to blind those measuring the main outcomes of the intervention?

[q]16: If any study results were based on data dredging, was this made clear?

[r]17: In trials and cohort studies, do the analysis adjust for different lengths of follow-up of participants, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?

[s]Not applicable.

[t]18: Were the statistical tests used to assess the main outcomes appropriate?

[u]19: Was intervention compliance reliable?

[v]20: Were the main outcome measures used accurate (valid and reliable)?

[w]21: Were the participants in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?

[x]22: Were the participants in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?

[y]23: Were participants randomized to the intervention groups?

[z]24: Was the randomized intervention assignment concealed from both participants and health care staff until recruitment was complete and irrevocable?

[aa]25: Was there adequate adjustment for confounding in the analysis from which the main findings were drawn?

[ab]26: Were losses to follow-up taken into account?

[ac]27: Did the study have sufficient power to detect a clinically important effect?

Quality Assessment of Measurement Scales

All of the outcome measures used in the 11 studies had the psychometric properties of the scale assessed previously with only the floor and ceiling effect of the BBT, JTHFT, the CAHAI, QuickDASH, and the UEFT not studied. The MAL had no evidence of content validity or predictive validity. The WMFT had no evidence of predictive and content validity or responsiveness.

From a clinimetric perspective, all of the scales with the exception of the UL item of the MAS, UEFT, and the JTHFT had defined MCID in the literature. All data, parametric or nonparametric, were analyzed using appropriate statistical methods. This quality assessment of outcome measurements used provides some confidence in the evidence reported by each study (see Multimedia Appendix 3 for further details of the outcome measures of quality assessment and Multimedia Appendix 6 for a summary of the included studies in this review).

The 11 included studies in this review had small sample sizes ranging from 5 to 99 participants at an average (mean) age of 57 years.

Discussion

Principal Findings

This review set out to answer the question What is the effectiveness of UL wearable technology for improving activity and participation in adult stroke survivors?

Following exclusions, outcome measure assessment and quality assessment, 11 studies were included (see Multimedia Appendix 6). Of the 11 studies included, only one [141] found significant between-group differences using the CAHAI. However, this study was assessed as being high risk (see Table 1 and Multimedia Appendix 4) because of having >20% dropout rate. This study also included acute stroke survivors <19 days poststroke, which could mean that improvements are subject to acute natural improvement such as the spontaneous recovery following stroke [152]. Some improvements were found across all the studies included for both the control and intervention groups who all had an increase in rehabilitation dosage. This may suggest that a key mechanism for improvement is increasing the amount of rehabilitation carried out, which has been recognized in the national clinical guidelines for stroke [5]. However, further research is required to distinguish between the mechanism of dosage and intensity where dosage is the amount of rehabilitation activity and intensity is the amount of rehabilitation over time [153]. In other words, is it more effective to carry out more rehabilitation or is more effective to carry out more rehabilitation in a shorter period of time?

The 11 studies included in this review had small sample sizes ranging from 5 to 99 participants at an average (mean) age of 57 years old, whereas in England, Wales, and Northern Ireland, the average age for men to have a stroke is 74 and the average age for women to have a stroke is 80 [154]. It is also important to note that only one of the RCTs [143] was assessed as being low risk using the CRoB tool and of the non-RCTs [148-151], all obtained low scores on the Downs and Black instrument (14-21 out of 32), particularly for the external validity and selection bias domains. This suggests that the results may be difficult to generalize to the wider stroke population. However, quality appraisal is reliant on adequate reporting and some interventions may rely heavily on direct clinical input, which negates the ability to blind the participants and assessors.

A total of 12 outcome measures were used across the studies (11 functionals and 1 participation) to assess the effectiveness of the intervention in their respective ICF domains. A review of each of the outcome measures used to determine if the psychometric properties, the clinimetric properties, and the method of analysis were suitable revealed that all data had been analyzed appropriately. However, the UL item of the MAS, UEFT, and the JTHFT are yet to establish the MCID. This is important as it represents the smallest improvement considered worthwhile by a patient. The concept of an MCID is offered as the new standard for determining effectiveness of a given treatment and describing patient satisfaction in reference to that treatment [155]. Although some studies may reveal statistical significance, this may not mean that the intervention has made a meaningful difference to the functional capability of the stroke survivor.

An RCT methodology aims to control the conditions of each arm of a study to reduce bias [156]. Using technology to facilitate independent poststroke rehabilitation involves combining complex interventions with a complex condition. No two strokes are the same and no two contexts of adoption are the same. There are many complex nuances involved in using a device, independently often in the home environment. Therefore, using methodologies that account for these differences are important such as realist evaluation [157,158].

The use of technology to facilitate independent poststroke rehabilitation has the potential to motivate stroke survivors in that they are often interactive, fun to use, and engaging [12]. Furthermore, technologies have the potential to measure intervention outcomes over long periods of time that would normally be undetectable (ie, muscle activity in microvolts); provide formative, summative, and real-time feedback to the user in ways that could not be provided by a therapist (ie, the use of readily available graphics); and provide guidance and instruction out of clinical sessions. However, the lack of large, robust clinical trials can limit the uptake and acceptance of technological interventions in clinical practice. Indeed, since Moore's law published in 1965, which observed that the number of transistors on integrated circuits doubles approximately every 2 years [159], one of the difficulties is keeping up with the speed of new technologies against the time it takes to provide high-level clinical evidence. However, the recent publication of the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies (DHTs) sets out to describe standards for the evidence that should be available, or developed, for DHTs to demonstrate their value in the UK health and care system could speed up the uptake of DHTs [160].

The results of the included studies were not combined for a meta-analysis due to the varied types of data collected for the primary outcome measures. It would also be difficult to compare primary outcomes across the included studies accurately as there were a wide variety of functional and participation outcome measures used across the studies.

Future research could focus on adopting the principles and concept of technology use rather than on a specific device. Nonetheless, conventional research rigor is still required including robust methodologies that account for the complexity of use, larger sample sizes that reflect the population, valid, reliable measurement tools with MCID values, and importantly, the technology is suitable for the purpose of use.

Conclusions

This review found that there is little evidence in the literature to support the use of wearable technologies to improve activity and participation for independent UL rehabilitation following a stroke. However, this may be because of small sample sizes and the limitations of using an RCT and randomized comparison trial methodology with a complex intervention and with a complex condition. The studies included in this review did highlight that improvements can be made when more rehabilitation is carried out, but the mechanisms of this are yet to be investigated fully. Future technologies may have the potential to measure outcomes, provide feedback, and engage users outside of clinical sessions. This could provide a platform for motivating stroke survivors to carry out more rehabilitation in the absence of a therapist, which could maximize recovery.

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 National Health Service, the NIHR, or the Department of Health, United Kingdom.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medline search strategy.
[DOCX File , 17 KB - jmir_v22i1e15981_app1.docx]

Multimedia Appendix 2
Full text papers and reasons for exclusions.
[DOCX File , 65 KB - jmir_v22i1e15981_app2.docx]

Multimedia Appendix 3
Details of outcome measure quality assessment.
[DOCX File , 14 KB - jmir_v22i1e15981_app3.docx]

Multimedia Appendix 4
Details of quality assessment for included RCTs.
[DOCX File , 15 KB - jmir_v22i1e15981_app4.docx]

Multimedia Appendix 5
Details of quality assessment for non-RCT study designs.
[DOCX File , 15 KB - jmir_v22i1e15981_app5.docx]

Multimedia Appendix 6
Summary of included studies in this review.
[DOCX File , 107 KB - jmir_v22i1e15981_app6.docx]

References

1. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):439-458 [FREE Full text] [doi: 10.1016/S1474-4422(19)30034-1] [Medline: 30871944]
2. Xu X, Vestesson E, Paley L, Desikan A, Wonderling D, Hoffman A, et al. The economic burden of stroke care in England, Wales and Northern Ireland: Using a national stroke register to estimate and report patient-level health economic outcomes in stroke. Eur Stroke J 2018 Mar;3(1):82-91 [FREE Full text] [doi: 10.1177/2396987317746516] [Medline: 29900412]
3. Kwah LK, Harvey LA, Diong J, Herbert RD. Models containing age and NIHSS predict recovery of ambulation and upper limb function six months after stroke: an observational study. J Physiother 2013 Sep;59(3):189-197 [FREE Full text] [doi: 10.1016/S1836-9553(13)70183-8] [Medline: 23896334]
4. Morris JH, van Wijck F, Joice S, Donaghy M. Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. Disabil Rehabil 2013 Feb;35(4):291-299. [doi: 10.3109/09638288.2012.691942] [Medline: 22691224]
5. Inercollegiate Stroke Working Party. National Clinical Guideline for Stroke. London: Royal College of Physicians; 2016.
6. Fryer CE, Luker JA, McDonnell MN, Hillier SL. Self management programmes for quality of life in people with stroke. Cochrane Database Syst Rev 2016 Aug 22(8):CD010442 [FREE Full text] [doi: 10.1002/14651858.CD010442.pub2] [Medline: 27545611]
7. Powell L, Parker J, Martyn St-James M, Mawson S. The effectiveness of lower-limb wearable technology for improving activity and participation in adult stroke survivors: a systematic review. J Med Internet Res 2016 Oct 7;18(10):e259 [FREE Full text] [doi: 10.2196/jmir.5891] [Medline: 27717920]
8. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil 2017 Mar 11;14(1):20 [FREE Full text] [doi: 10.1186/s12984-017-0229-y] [Medline: 28284228]
9. Parker J, Mountain G, Hammerton J. A review of the evidence underpinning the use of visual and auditory feedback for computer technology in post-stroke upper-limb rehabilitation. Disabil Rehabil Assist Technol 2011;6(6):465-472. [doi: 10.3109/17483107.2011.556209] [Medline: 21314295]
10. Burridge JH, Lee AC, Turk R, Stokes M, Whitall J, Vaidyanathan R, et al. Telehealth, wearable sensors, and the internet: will they improve stroke outcomes through increased intensity of therapy, motivation, and adherence to rehabilitation programs? J Neurol Phys Ther 2017 Jul;41(Suppl 3):S32-S38. [doi: 10.1097/NPT.0000000000000183] [Medline: 28628594]
11. Morone G, Giardi S, Gooshchy S, Iosa M, Paolucci S. Wearable Devices and Virtual Reality for Neurorehabilitation: An Opportunity for Home Rehabilitation. In: Proceedings of the 2018 International Conference on NeuroRehabilitation. 2018 Presented at: ICNR'18; October 16-20, 2018; Pisa, Italy. [doi: 10.1007/978-3-030-01845-0_120]

12. Rodgers MM, Alon G, Pai VM, Conroy RS. Wearable technologies for active living and rehabilitation: current research challenges and future opportunities. J Rehabil Assist Technol Eng 2019;6:2055668319839607 [FREE Full text] [doi: 10.1177/2055668319839607] [Medline: 31245033]
13. Dobkin BH, Martinez C. Wearable sensors to monitor, enable feedback, and measure outcomes of activity and practice. Curr Neurol Neurosci Rep 2018 Oct 6;18(12):87. [doi: 10.1007/s11910-018-0896-5] [Medline: 30293160]
14. Gebruers N, Vanroy C, Truijen S, Engelborghs S, de Deyn PP. Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures. Arch Phys Med Rehabil 2010 Feb;91(2):288-297. [doi: 10.1016/j.apmr.2009.10.025] [Medline: 20159136]
15. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
16. Steins D, Dawes H, Esser P, Collett J. Wearable accelerometry-based technology capable of assessing functional activities in neurological populations in community settings: a systematic review. J Neuroeng Rehabil 2014 Mar 13;11:36 [FREE Full text] [doi: 10.1186/1743-0003-11-36] [Medline: 24625308]
17. Zheng H, Black ND, Harris ND. Position-sensing technologies for movement analysis in stroke rehabilitation. Med Biol Eng Comput 2005 Jul;43(4):413-420. [doi: 10.1007/bf02344720] [Medline: 16255421]
18. Wang Q, Timmermans A, Chen W, Jia J, Ding L, Xiong L, et al. Stroke patients' acceptance of a smart garment for supporting upper extremity rehabilitation. IEEE J Transl Eng Health Med 2018;6:2101009 [FREE Full text] [doi: 10.1109/JTEHM.2018.2853549] [Medline: 30519515]
19. Davies RJ, Parker J, McCullagh P, Zheng H, Nugent C, Black ND, et al. A personalized self-management rehabilitation system for stroke survivors: a quantitative gait analysis using a smart insole. JMIR Rehabil Assist Technol 2016 Nov 8;3(2):e11 [FREE Full text] [doi: 10.2196/rehab.5449] [Medline: 28582260]
20. 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]
21. Munoz-Organero M, Parker J, Powell L, Davies R, Mawson S. Sensor optimization in smart insoles for post-stroke gait asymmetries using total variation and L 1 distances. IEEE Sensors J 2017;17(10):3142-3151. [doi: 10.1109/jsen.2017.2686641]
22. Munoz-Organero M, Parker J, Powell L, Mawson S. Assessing walking strategies using insole pressure sensors for stroke survivors. Sensors (Basel) 2016 Oct 1;16(10):pii: E1631 [FREE Full text] [doi: 10.3390/s16101631] [Medline: 27706077]
23. Asada H, Shaltis P, Reisner A, Rhee S, Hutchinson R. Mobile monitoring with wearable photoplethysmographic biosensors. IEEE Eng Med Biol Mag 2003;22(3):28-40. [doi: 10.1109/memb.2003.1213624] [Medline: 12845817]
24. Shaltis PA, Reisner A, Asada HH. Wearable, cuff-less PPG-based blood pressure monitor with novel height sensor. Conf Proc IEEE Eng Med Biol Soc 2006;1:908-911. [doi: 10.1109/IEMBS.2006.260027] [Medline: 17946006]
25. Veerbeek JM, Langbroek-Amersfoort AC, van Wegen EE, Meskers CG, Kwakkel G. Effects of robot-assisted therapy for the upper limb after stroke. Neurorehabil Neural Repair 2017 Feb;31(2):107-121. [doi: 10.1177/1545968316666957] [Medline: 27597165]
26. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
27. Eraifej J, Clark W, France B, Desando S, Moore D. Effectiveness of upper limb functional electrical stimulation after stroke for the improvement of activities of daily living and motor function: a systematic review and meta-analysis. Syst Rev 2017 Feb 28;6(1):40 [FREE Full text] [doi: 10.1186/s13643-017-0435-5] [Medline: 28245858]
28. Howlett OA, Lannin NA, Ada L, McKinstry C. Functional electrical stimulation improves activity after stroke: a systematic review with meta-analysis. Arch Phys Med Rehabil 2015 May;96(5):934-943. [doi: 10.1016/j.apmr.2015.01.013] [Medline: 25634620]
29. Teasell RW, Bhogal SK, Foley NC, Speechley MR. Gait retraining post stroke. Top Stroke Rehabil 2003;10(2):34-65. [doi: 10.1310/UDXE-MJFF-53V2-EAP0] [Medline: 13680517]
30. Mahmood A, Veluswamy SK, Hombali A, Mullick A, Solomon JM. Effect of transcutaneous electrical nerve stimulation on spasticity in adults with stroke: a systematic review and meta-analysis. Arch Phys Med Rehabil 2019 Apr;100(4):751-768. [doi: 10.1016/j.apmr.2018.10.016] [Medline: 30452892]
31. Marcolino MA, Hauck M, Stein C, Schardong J, Pagnussat AD, Plentz RD. Effects of transcutaneous electrical nerve stimulation alone or as additional therapy on chronic post-stroke spasticity: systematic review and meta-analysis of randomized controlled trials. Disabil Rehabil 2018 Oct 16:1-13. [doi: 10.1080/09638288.2018.1503736] [Medline: 30326752]
32. Mills PB, Dossa F. Transcutaneous electrical nerve stimulation for management of limb spasticity: a systematic review. Am J Phys Med Rehabil 2016 Apr;95(4):309-318. [doi: 10.1097/PHM.0000000000000437] [Medline: 26829077]
33. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
34. World Health Organisation. World Health Organization. 2001. International Classification of Functioning, Disability and Health. URL: https://www.who.int/classifications/icf/en/ [accessed 2019-11-07]

35. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, et al. ICF Core Sets for stroke. J Rehabil Med 2004 Jul(44 Suppl):135-141 [FREE Full text] [doi: 10.1080/16501960410016776] [Medline: 15370761]
36. Gottlieb A, Golander H, Bar-Tal Y, Gottlieb D. The influence of social support and perceived control on handicap and quality of life after stroke. Aging (Milano) 2001 Feb;13(1):11-15. [doi: 10.1007/bf03351488] [Medline: 11292146]
37. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
38. Godfrey A, Hetherington V, Shum H, Bonato P, Lovell NH, Stuart S. From A to Z: Wearable technology explained. Maturitas 2018 Jul;113:40-47 [FREE Full text] [doi: 10.1016/j.maturitas.2018.04.012] [Medline: 29903647]
39. McKenzie JE, Brennan SE, Ryan RE, Thompson J, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventionsversion 6.0. London: Cochrane; 2019.
40. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. Disabil Rehabil 2005 Feb 18;27(4):191-207. [doi: 10.1080/09638280400008537] [Medline: 15824050]
41. Platz T, Pinkowski C, van Wijck F, Kim I, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil 2005 Jun;19(4):404-411. [doi: 10.1191/0269215505cr832oa] [Medline: 15929509]
42. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res 1981;4(4):483-492. [doi: 10.1097/00004356-198112000-00001] [Medline: 7333761]
43. Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. J Formos Med Assoc 2001 Aug;100(8):526-532. [Medline: 11678002]
44. Barreca S, Gowland CK, Stratford P, Huijbregts M, Griffiths J, Torresin W, et al. Development of the Chedoke arm and hand activity inventory: theoretical constructs, item generation, and selection. Top Stroke Rehabil 2004;11(4):31-42. [doi: 10.1310/JU8P-UVK6-68VW-CF3W] [Medline: 15592988]
45. Barreca SR, Stratford PW, Lambert CL, Masters LM, Streiner DL. Test-retest reliability, validity, and sensitivity of the Chedoke arm and hand activity inventory: a new measure of upper-limb function for survivors of stroke. Arch Phys Med Rehabil 2005 Aug;86(8):1616-1622. [doi: 10.1016/j.apmr.2005.03.017] [Medline: 16084816]
46. Barreca SR, Stratford PW, Masters LM, Lambert CL, Griffiths J. Comparing 2 versions of the Chedoke Arm and Hand Activity Inventory with the Action Research Arm Test. Phys Ther 2006 Feb;86(2):245-253. [Medline: 16445338]
47. Barreca SR, Stratford PW, Masters LM, Lambert CL, Griffiths J, McBay C. Validation of three shortened versions of the Chedoke arm and Hand Activity Inventory. Physiother Can 2006;58(02):148 [FREE Full text] [doi: 10.2310/6640.2006.00031]
48. Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. Arch Phys Med Rehabil 1969;50(6):311-319. [Medline: 5788487]
49. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke 2001 Jul;32(7):1635-1639. [doi: 10.1161/01.str.32.7.1635] [Medline: 11441212]
50. Uswatte G, Taub E, Morris D, Light K, Thompson PA. The Motor Activity Log-28: assessing daily use of the hemiparetic arm after stroke. Neurology 2006 Oct 10;67(7):1189-1194. [doi: 10.1212/01.wnl.0000238164.90657.c2] [Medline: 17030751]
51. Poole JL, Whitney SL. Motor assessment scale for stroke patients: concurrent validity and interrater reliability. Arch Phys Med Rehabil 1988 Mar;69(3 Pt 1):195-197. [Medline: 3348720]
52. Duncan PW, Bode RK, Min Lai S, Perera S, Glycine Antagonist in Neuroprotection Americans Investigators. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. Arch Phys Med Rehabil 2003 Jul;84(7):950-963. [doi: 10.1016/s0003-9993(03)00035-2] [Medline: 12881816]
53. Adams SA, Ashburn A, Pickering RM, Taylor D. The scalability of the Rivermead Motor Assessment in acute stroke patients. Clin Rehabil 1997 Feb;11(1):42-51. [doi: 10.1177/026921559701100107] [Medline: 9065359]
54. Carroll D. A quantitative test of upper extremity function. J Chronic Dis 1965 May;18:479-491. [doi: 10.1016/0021-9681(65)90030-5] [Medline: 14293031]
55. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005 May;87(5):1038-1046. [doi: 10.2106/JBJS.D.02060] [Medline: 15866967]
56. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair 2002 Sep;16(3):232-240. [doi: 10.1177/154596802401105171] [Medline: 12234086]
57. O'Dell MW, Kim G, Rivera L, Fieo R, Christos P, Polistena C, et al. A psychometric evaluation of the Arm Motor Ability Test. J Rehabil Med 2013 Jun;45(6):519-527 [FREE Full text] [doi: 10.2340/16501977-1138] [Medline: 23532275]
58. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983 Sep;17(1):45-56. [doi: 10.1016/0304-3959(83)90126-4] [Medline: 6226917]
59. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey: Wiley; 2008:187-241.

60. Tanaka N, Saitou H, Takao T, Iizuka N, Okuno J, Yano H, et al. Effects of gait rehabilitation with a footpad-type locomotion interface in patients with chronic post-stroke hemiparesis: a pilot study. Clin Rehabil 2012 Aug;26(8):686-695. [doi: 10.1177/0269215511432356] [Medline: 22275464]
61. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
62. Tyson S, Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. Clin Rehabil 2009 Nov;23(11):1018-1033. [doi: 10.1177/0269215509339004] [Medline: 19786420]
63. Altaim T. White Rose eTheses Online. 2015. Developing a Functional Outcome Measure for Individuals With Low Back Pain Within a Jordanian Physiotherapy Service. URL: http://etheses.whiterose.ac.uk/9602/1/Thesis_Thamer%20Altaim_Draft_8August2015_Final.pdf [accessed 2019-11-07]
64. de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. J Clin Epidemiol 2003 Dec;56(12):1137-1141. [doi: 10.1016/j.jclinepi.2003.08.012] [Medline: 14680660]
65. Feinstein AR. An additional basic science for clinical medicine: IV. The development of clinimetrics. Ann Intern Med 1983 Dec;99(6):843-848. [doi: 10.7326/0003-4819-99-6-843] [Medline: 6651026]
66. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38(1):27-36. [doi: 10.1016/0021-9681(85)90005-0] [Medline: 3972947]
67. Okasheh RO. Sheffield Hallam University Research Archive. 2011. Clinical Measurement of Functional Outcomes of Pulmonary Rehabilitation. URL: http://shura.shu.ac.uk/20771/1/10702873.pdf [accessed 2019-11-07]
68. Mawson SJ. Measuring physiotherapy outcome in stroke rehabilitation. Physiother 1993;79(11):762-765. [doi: 10.1016/s0031-9406(10)60057-0]
69. Michels E. Measurement in physical therapy. On the rules for assigning numerals to observations. Phys Ther 1983 Feb;63(2):209-215. [doi: 10.1093/ptj/63.2.209] [Medline: 6823471]
70. Stevens SS. On the theory of scales of measurement. Science 1946 Jun 7;103(2684):677-680. [doi: 10.1126/science.103.2684.677] [Medline: 17750512]
71. Weiss PL, Rand D, Katz N, Kizony R. Video capture virtual reality as a flexible and effective rehabilitation tool. J Neuroeng Rehabil 2004 Dec 20;1(1):12 [FREE Full text] [doi: 10.1186/1743-0003-1-12] [Medline: 15679949]
72. Nelson AM. Addressing the threat of evidence-based practice to qualitative inquiry through increasing attention to quality: a discussion paper. Int J Nurs Stud 2008 Feb;45(2):316-322. [doi: 10.1016/j.ijnurstu.2007.01.012] [Medline: 17362956]
73. Mullen EJ, Streiner DL. The evidence for and against evidence-based practice. In: Roberts AR, Yeager K, editors. Foundations of Evidence-Based Social Work Practice. Oxford, UK: Oxford University Press; 2006:21-34.
74. Ang KK, Guan C, Phua KS, Wang C, Zhao L, Teo WP, et al. Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation. Arch Phys Med Rehabil 2015 Mar;96(3 Suppl):S79-S87. [doi: 10.1016/j.apmr.2014.08.008] [Medline: 25721551]
75. Bach P, Knoblich G, Gunter TC, Friederici AD, Prinz W. Action comprehension: deriving spatial and functional relations. J Exp Psychol Hum Percept Perform 2005 Jun;31(3):465-479. [doi: 10.1037/0096-1523.31.3.465] [Medline: 15982126]
76. Cameirão MS, Badia SB, Duarte E, Frisoli A, Verschure PF. The combined impact of virtual reality neurorehabilitation and its interfaces on upper extremity functional recovery in patients with chronic stroke. Stroke 2012 Oct;43(10):2720-2728. [doi: 10.1161/STROKEAHA.112.653196] [Medline: 22871683]
77. Carey JR, Kimberley TJ, Lewis SM, Auerbach EJ, Dorsey L, Rundquist P, et al. Analysis of fMRI and finger tracking training in subjects with chronic stroke. Brain 2002 Apr;125(Pt 4):773-788. [doi: 10.1093/brain/awf091] [Medline: 11912111]
78. Germanotta M, Cruciani A, Pecchioli C, Loreti S, Spedicato A, Meotti M, et al. Reliability, validity and discriminant ability of the instrumental indices provided by a novel planar robotic device for upper limb rehabilitation. J Neuroeng Rehabil 2018 May 16;15(1):39 [FREE Full text] [doi: 10.1186/s12984-018-0385-8] [Medline: 29769127]
79. Hayward KS, Barker RN, Brauer SG, Lloyd D, Horsley SA, Carson RG. SMART Arm with outcome-triggered electrical stimulation: a pilot randomized clinical trial. Top Stroke Rehabil 2013;20(4):289-298. [doi: 10.1310/tsr2004-289] [Medline: 23893828]
80. Hussain I, Salvietti G, Spagnoletti G, Malvezzi M, Cioncoloni D, Rossi S, et al. A soft supernumerary robotic finger and mobile arm support for grasping compensation and hemiparetic upper limb rehabilitation. Robot Auton Syst 2017;93:1-12 [FREE Full text] [doi: 10.1016/j.robot.2017.03.015]
81. Kairy D, Veras M, Archambault P, Hernandez A, Higgins J, Levin MF, et al. Maximizing post-stroke upper limb rehabilitation using a novel telerehabilitation interactive virtual reality system in the patient's home: study protocol of a randomized clinical trial. Contemp Clin Trials 2016 Mar;47:49-53. [doi: 10.1016/j.cct.2015.12.006] [Medline: 26655433]
82. Kim T, Kim S, Lee B. Effects of action observational training plus brain-computer interface-based functional electrical stimulation on paretic arm motor recovery in patient with stroke: a randomized controlled trial. Occup Ther Int 2016 Mar;23(1):39-47. [doi: 10.1002/oti.1403] [Medline: 26301519]

83. Kim GJ, Hinojosa J, Rao AK, Batavia M, O'Dell MW. Randomized trial on the effects of attentional focus on motor training of the upper extremity using robotics with individuals after chronic stroke. Arch Phys Med Rehabil 2017 Oct;98(10):1924-1931. [doi: 10.1016/j.apmr.2017.06.005] [Medline: 28652064]
84. Kimberley TJ, Lewis SM, Auerbach EJ, Dorsey LL, Lojovich JM, Carey JR. Electrical stimulation driving functional improvements and cortical changes in subjects with stroke. Exp Brain Res 2004 Feb;154(4):450-460. [doi: 10.1007/s00221-003-1695-y] [Medline: 14618287]
85. Kiper P, Agostini M, Luque-Moreno C, Tonin P, Turolla A. Reinforced feedback in virtual environment for rehabilitation of upper extremity dysfunction after stroke: preliminary data from a randomized controlled trial. Biomed Res Int 2014;2014:752128 [FREE Full text] [doi: 10.1155/2014/752128] [Medline: 24745024]
86. Klamroth-Marganska V, Blanco J, Campen K, Curt A, Dietz V, Ettlin T, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. Lancet Neurol 2014 Feb;13(2):159-166. [doi: 10.1016/S1474-4422(13)70305-3] [Medline: 24382580]
87. Lee JM, Jung J, Park KW, Shin E, Oh SK, Bae J, et al. Harmonizing Optimal Strategy for Treatment of coronary artery diseases--comparison of REDUCtion of prasugrEl dose or POLYmer TECHnology in ACS patients (HOST-REDUCE-POLYTECH-ACS RCT): study protocol for a randomized controlled trial. Trials 2015 Sep 15;16:409 [FREE Full text] [doi: 10.1186/s13063-015-0925-5] [Medline: 26374625]
88. Lemmens RJ, Timmermans AA, Janssen-Potten YJ, Pulles SA, Geers RP, Bakx WG, et al. Accelerometry measuring the outcome of robot-supported upper limb training in chronic stroke: a randomized controlled trial. PLoS One 2014;9(5):e96414 [FREE Full text] [doi: 10.1371/journal.pone.0096414] [Medline: 24823925]
89. Leonardis D, Barsotti M, Loconsole C, Solazzi M, Troncossi M, Mazzotti C, et al. An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. IEEE Trans Haptics 2015;8(2):140-151. [doi: 10.1109/TOH.2015.2417570] [Medline: 25838528]
90. Li H, Huang G, Lin Q, Zhao J, Lo WA, Mao Y, et al. Combining movement-related cortical potentials and event-related desynchronization to study movement preparation and execution. Front Neurol 2018;9:822 [FREE Full text] [doi: 10.3389/fneur.2018.00822] [Medline: 30344504]
91. Linder SM, Rosenfeldt AB, Reiss A, Buchanan S, Sahu K, Bay CR, et al. The home stroke rehabilitation and monitoring system trial: a randomized controlled trial. Int J Stroke 2013 Jan;8(1):46-53 [FREE Full text] [doi: 10.1111/j.1747-4949.2012.00971.x] [Medline: 23280269]
92. Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. N Engl J Med 2010 May 13;362(19):1772-1783 [FREE Full text] [doi: 10.1056/NEJMoa0911341] [Medline: 20400552]
93. Lum PS, Burgar CG, Shor PC, Majmundar M, Van der Loos M. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. Arch Phys Med Rehabil 2002 Jul;83(7):952-959. [doi: 10.1053/apmr.2001.33101] [Medline: 12098155]
94. Lum PS, Burgar CG, Van der Loos M, Shor PC, Majmundar M, Yap R. MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: a follow-up study. J Rehabil Res Dev 2006;43(5):631-642 [FREE Full text] [doi: 10.1682/jrrd.2005.02.0044] [Medline: 17123204]
95. Masiero S, Celia A, Armani M, Rosati G. A novel robot device in rehabilitation of post-stroke hemiplegic upper limbs. Aging Clin Exp Res 2006 Dec;18(6):531-535. [Medline: 17255643]
96. Masiero S, Celia A, Rosati G, Armani M. Robotic-assisted rehabilitation of the upper limb after acute stroke. Arch Phys Med Rehabil 2007 Feb;88(2):142-149. [doi: 10.1016/j.apmr.2006.10.032] [Medline: 17270510]
97. Masiero S, Armani M, Rosati G. Upper-limb robot-assisted therapy in rehabilitation of acute stroke patients: focused review and results of new randomized controlled trial. J Rehabil Res Dev 2011;48(4):355-366 [FREE Full text] [doi: 10.1682/jrrd.2010.04.0063] [Medline: 21674388]
98. Masiero S, Armani M, Ferlini G, Rosati G, Rossi A. Randomized trial of a robotic assistive device for the upper extremity during early inpatient stroke rehabilitation. Neurorehabil Neural Repair 2014 May;28(4):377-386. [doi: 10.1177/1545968313513073] [Medline: 24316679]
99. McCabe J, Monkiewicz M, Holcomb J, Pundik S, Daly JJ. Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Jun;96(6):981-990 [FREE Full text] [doi: 10.1016/j.apmr.2014.10.022] [Medline: 25461822]
100. McNulty PA, Thompson-Butel AG, Faux SG, Lin G, Katrak PH, Harris LR, et al. The efficacy of Wii-based Movement Therapy for upper limb rehabilitation in the chronic poststroke period: a randomized controlled trial. Int J Stroke 2015 Dec;10(8):1253-1260. [doi: 10.1111/ijs.12594] [Medline: 26332338]
101. Metzger J, Lambercy O, Califfi A, Dinacci D, Petrillo C, Rossi P, et al. Assessment-driven selection and adaptation of exercise difficulty in robot-assisted therapy: a pilot study with a hand rehabilitation robot. J Neuroeng Rehabil 2014 Nov 15;11:154 [FREE Full text] [doi: 10.1186/1743-0003-11-154] [Medline: 25399249]
102. Orihuela-Espina F, Roldán GF, Sánchez-Villavicencio I, Palafox L, Leder R, Sucar LE, et al. Robot training for hand motor recovery in subacute stroke patients: a randomized controlled trial. J Hand Ther 2016;29(1):51-7; quiz 57. [doi: 10.1016/j.jht.2015.11.006] [Medline: 26847320]

103. Prange GB, Kottink AI, Buurke JH, Eckhardt MM, van Keulen-Rouweler BJ, Ribbers GM, et al. The effect of arm support combined with rehabilitation games on upper-extremity function in subacute stroke: a randomized controlled trial. Neurorehabil Neural Repair 2015 Feb;29(2):174-182. [doi: 10.1177/1545968314535985] [Medline: 24878589]
104. Piron L, Turolla A, Agostini M, Zucconi C, Cortese F, Zampolini M, et al. Exercises for paretic upper limb after stroke: a combined virtual-reality and telemedicine approach. J Rehabil Med 2009 Nov;41(12):1016-1102 [FREE Full text] [doi: 10.2340/16501977-0459] [Medline: 19841835]
105. Rand D, Givon N, Weingarden H, Nota A, Zeilig G. Eliciting upper extremity purposeful movements using video games: a comparison with traditional therapy for stroke rehabilitation. Neurorehabil Neural Repair 2014 Oct;28(8):733-739. [doi: 10.1177/1545968314521008] [Medline: 24515927]
106. Sale P, Mazzoleni S, Lombardi V, Galafate D, Massimiani MP, Posteraro F, et al. Recovery of hand function with robot-assisted therapy in acute stroke patients. Int J Rehabil Res 2014;37(3):236-242. [doi: 10.1097/mrr.0000000000000059]
107. Saposnik G, Mamdani M, Bayley M, Thorpe KE, Hall J, Cohen LG, EVREST Steering Committee, EVREST Study Group for the Stroke Outcome Research Canada Working Group. Effectiveness of Virtual Reality Exercises in STroke Rehabilitation (EVREST): rationale, design, and protocol of a pilot randomized clinical trial assessing the Wii gaming system. Int J Stroke 2010 Feb;5(1):47-51 [FREE Full text] [doi: 10.1111/j.1747-4949.2009.00404.x] [Medline: 20088994]
108. Saposnik G, Cohen LG, Mamdani M, Pooyania S, Ploughman M, Cheung D, Stroke Outcomes Research Canada. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. Lancet Neurol 2016 Sep;15(10):1019-1027 [FREE Full text] [doi: 10.1016/S1474-4422(16)30121-1] [Medline: 27365261]
109. Takahashi K, Domen K, Sakamoto T, Toshima M, Otaka Y, Seto M, et al. Efficacy of upper extremity robotic therapy in subacute poststroke hemiplegia: an exploratory randomized trial. Stroke 2016 May;47(5):1385-1388. [doi: 10.1161/STROKEAHA.115.012520] [Medline: 27006452]
110. Timmermans AA, Lemmens RJ, Monfrance M, Geers RP, Bakx W, Smeets RJ, et al. Effects of task-oriented robot training on arm function, activity, and quality of life in chronic stroke patients: a randomized controlled trial. J Neuroeng Rehabil 2014 Mar 31;11:45 [FREE Full text] [doi: 10.1186/1743-0003-11-45] [Medline: 24684808]
111. Tomić TJ, Savić AM, Vidaković AS, Rodić SZ, Isaković MS, Rodríguez-de-Pablo C, et al. ArmAssist robotic system versus matched conventional therapy for poststroke upper limb rehabilitation: a randomized clinical trial. Biomed Res Int 2017;2017:7659893 [FREE Full text] [doi: 10.1155/2017/7659893] [Medline: 28251157]
112. Türkbey TA, Kutlay S, Gök H. Clinical feasibility of Xbox KinectTM training for stroke rehabilitation: a single-blind randomized controlled pilot study. J Rehabil Med 2017 Jan 19;49(1):22-29 [FREE Full text] [doi: 10.2340/16501977-2183] [Medline: 27973678]
113. Vanoglio F, Bernocchi P, Mulè C, Garofali F, Mora C, Taveggia G, et al. Feasibility and efficacy of a robotic device for hand rehabilitation in hemiplegic stroke patients: a randomized pilot controlled study. Clin Rehabil 2017 Mar;31(3):351-360. [doi: 10.1177/0269215516642606] [Medline: 27056250]
114. Volpe BT, Lynch D, Rykman-Berland A, Ferraro M, Galgano M, Hogan N, et al. Intensive sensorimotor arm training mediated by therapist or robot improves hemiparesis in patients with chronic stroke. Neurorehabil Neural Repair 2008;22(3):305-310 [FREE Full text] [doi: 10.1177/1545968307311102] [Medline: 18184932]
115. Wei XJ, Tong KY, Hu XL. The responsiveness and correlation between Fugl-Meyer Assessment, Motor Status Scale, and the Action Research Arm Test in chronic stroke with upper-extremity rehabilitation robotic training. Int J Rehabil Res 2011 Dec;34(4):349-356. [doi: 10.1097/MRR.0b013e32834d330a] [Medline: 22044987]
116. Ang KK, Guan C, Chua K, Ang BT, Kuah C, Wang C, et al. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. Conf Proc IEEE Eng Med Biol Soc 2010;2010:5549-5552. [doi: 10.1109/IEMBS.2010.5626782] [Medline: 21096475]
117. Barker RN, Hayward KS, Carson RG, Lloyd D, Brauer SG. SMART arm training with outcome-triggered electrical stimulation in subacute stroke survivors with severe arm disability: a randomized controlled trial. Neurorehabil Neural Repair 2017 Dec;31(12):1005-1016. [doi: 10.1177/1545968317744276] [Medline: 29199503]
118. Huang Y, Lai WP, Qian Q, Hu X, Tam EW, Zheng Y. Translation of robot-assisted rehabilitation to clinical service: a comparison of the rehabilitation effectiveness of EMG-driven robot hand assisted upper limb training in practical clinical service and in clinical trial with laboratory configuration for chronic stroke. Biomed Eng Online 2018 Jun 25;17(1):91 [FREE Full text] [doi: 10.1186/s12938-018-0516-2] [Medline: 29941043]
119. Hwang CH, Seong JW, Son D. Individual finger synchronized robot-assisted hand rehabilitation in subacute to chronic stroke: a prospective randomized clinical trial of efficacy. Clin Rehabil 2012 Aug;26(8):696-704. [doi: 10.1177/0269215511431473] [Medline: 22261813]
120. Kim H, Miller LM, Fedulow I, Simkins M, Abrams GM, Byl N, et al. Kinematic data analysis for post-stroke patients following bilateral versus unilateral rehabilitation with an upper limb wearable robotic system. IEEE Trans Neural Syst Rehabil Eng 2013 Mar;21(2):153-164. [doi: 10.1109/TNSRE.2012.2207462] [Medline: 22855233]
121. Liao W, Wu C, Hsieh Y, Lin K, Chang W. Effects of robot-assisted upper limb rehabilitation on daily function and real-world arm activity in patients with chronic stroke: a randomized controlled trial. Clin Rehabil 2012 Feb;26(2):111-120. [doi: 10.1177/0269215511416383] [Medline: 21840917]

122. Page SJ, Hill V, White S. Portable upper extremity robotics is as efficacious as upper extremity rehabilitative therapy: a randomized controlled pilot trial. Clin Rehabil 2013 Jun;27(6):494-503 [FREE Full text] [doi: 10.1177/0269215512464795] [Medline: 23147552]
123. Susanto EA, Tong RK, Ockenfeld C, Ho NS. Efficacy of robot-assisted fingers training in chronic stroke survivors: a pilot randomized-controlled trial. J Neuroeng Rehabil 2015 Apr 25;12:42 [FREE Full text] [doi: 10.1186/s12984-015-0033-5] [Medline: 25906983]
124. Kumar D, Gubbi J, Yan B, Palaniswami M. Motor recovery monitoring in post acute stroke patients using wireless accelerometer and cross-correlation. Conf Proc IEEE Eng Med Biol Soc 2013;2013:6703-6706. [doi: 10.1109/EMBC.2013.6611094] [Medline: 24111281]
125. Bergmann J, Krewer C, Bauer, Koenig A, Riener R, Müller F. Virtual reality to augment robot-assisted gait training in non-ambulatory patients with a subacute stroke: a pilot randomized controlled trial. Eur J Phys Rehabil Med 2018 Jun;54(3):397-407 [FREE Full text] [doi: 10.23736/S1973-9087.17.04735-9] [Medline: 29265791]
126. Cannell J, Jovic E, Rathjen A, Lane K, Tyson AM, Callisaya ML, et al. The efficacy of interactive, motion capture-based rehabilitation on functional outcomes in an inpatient stroke population: a randomized controlled trial. Clin Rehabil 2018 Feb;32(2):191-200 [FREE Full text] [doi: 10.1177/0269215517720790] [Medline: 28719977]
127. Cruz VT, Bento V, Ruano L, Ribeiro DD, Fontão L, Mateus C, et al. Motor task performance under vibratory feedback early poststroke: single center, randomized, cross-over, controlled clinical trial. Sci Rep 2014 Jul 11;4:5670 [FREE Full text] [doi: 10.1038/srep05670] [Medline: 25011667]
128. Lin L, Lin Y, Lin Z, Chuang L, Hsu W, Lin Y. Feasibility and efficacy of wearable devices for upper limb rehabilitation in patients with chronic stroke: a randomized controlled pilot study. Eur J Phys Rehabil Med 2018 Jun;54(3):388-396 [FREE Full text] [doi: 10.23736/S1973-9087.17.04691-3] [Medline: 28627862]
129. Shimodozono M, Noma T, Matsumoto S, Miyata R, Etoh S, Kawahira K. Repetitive facilitative exercise under continuous electrical stimulation for severe arm impairment after sub-acute stroke: a randomized controlled pilot study. Brain Inj 2014;28(2):203-210. [doi: 10.3109/02699052.2013.860472] [Medline: 24304090]
130. Sullivan JE, Hurley D, Hedman LD. Afferent stimulation provided by glove electrode during task-specific arm exercise following stroke. Clin Rehabil 2012 Nov;26(11):1010-1020. [doi: 10.1177/0269215512442915] [Medline: 22561099]
131. Curado MR, Cossio EG, Broetz D, Agostini M, Cho W, Brasil FL, et al. Residual upper arm motor function primes innervation of paretic forearm muscles in chronic stroke after brain-machine interface (BMI) training. PLoS One 2015;10(10):e0140161 [FREE Full text] [doi: 10.1371/journal.pone.0140161] [Medline: 26495971]
132. Lee MM, Cho H, Song CH. The mirror therapy program enhances upper-limb motor recovery and motor function in acute stroke patients. Am J Phys Med Rehabil 2012 Aug;91(8):689-96, quiz 697. [doi: 10.1097/PHM.0b013e31824fa86d] [Medline: 22469877]
133. Chen H, Chen CC, Hsueh I, Huang S, Hsieh C. Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. Neurorehabil Neural Repair 2009 Jun;23(5):435-440. [doi: 10.1177/1545968308331146] [Medline: 19261767]
134. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Arch Phys Med Rehabil 2008 Sep;89(9):1693-1700 [FREE Full text] [doi: 10.1016/j.apmr.2008.02.022] [Medline: 18760153]
135. Hsieh Y, Wang C, Wu S, Chen P, Sheu C, Hsieh C. Establishing the minimal clinically important difference of the Barthel Index in stroke patients. Neurorehabil Neural Repair 2007;21(3):233-238. [doi: 10.1177/1545968306294729] [Medline: 17351082]
136. AbilityLab Home | Shirley Ryan AbilityLab. 2019. Rivermead Motor Assessment. URL: https://www.sralab.org/rehabilitation-measures/rivermead-motor-assessment [accessed 2019-09-23]
137. AbilityLab Home | Shirley Ryan AbilityLab. 2019. Chedoke Arm and Hand Activity Inventory. URL: https://www.sralab.org/rehabilitation-measures/chedoke-arm-and-hand-activity-inventory-7 [accessed 2019-09-23]
138. Dixon D, Johnston M, McQueen M, Court-Brown C. The Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH) can measure the impairment, activity limitations and participation restriction constructs from the International Classification of Functioning, Disability and Health (ICF). BMC Musculoskelet Disord 2008 Aug 20;9:114 [FREE Full text] [doi: 10.1186/1471-2474-9-114] [Medline: 18715495]
139. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther 2014 Jan;44(1):30-39. [doi: 10.2519/jospt.2014.4893] [Medline: 24175606]
140. Fulk GD, Ludwig M, Dunning K, Golden S, Boyne P, West T. How much change in the stroke impact scale-16 is important to people who have experienced a stroke? Top Stroke Rehabil 2010;17(6):477-483. [doi: 10.1310/tsr1706-477] [Medline: 21239371]
141. Cameirão M, Badia S, Duarte E, Verschure PF. Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: a randomized controlled pilot study in the acute phase of stroke using the rehabilitation gaming system. Restor Neurol Neurosci 2011;29(5):287-298. [doi: 10.3233/RNN-2011-0599] [Medline: 21697589]

142. Nijenhuis SM, Prange-Lasonder GB, Stienen AH, Rietman JS, Buurke JH. Effects of training with a passive hand orthosis and games at home in chronic stroke: a pilot randomised controlled trial. Clin Rehabil 2017 Feb;31(2):207-216. [doi: 10.1177/0269215516629722] [Medline: 26869596]
143. Lannin NA, Cusick A, Hills C, Kinnear B, Vogel K, Matthews K, et al. Upper limb motor training using a Saebo orthosis is feasible for increasing task-specific practice in hospital after stroke. Aust Occup Ther J 2016 Dec;63(6):364-372. [doi: 10.1111/1440-1630.12330] [Medline: 27646624]
144. Wolf SL, Sahu K, Bay RC, Buchanan S, Reiss A, Linder S, et al. The HAAPI (Home Arm Assistance Progression Initiative) Trial: a novel robotics delivery approach in stroke rehabilitation. Neurorehabil Neural Repair 2015;29(10):958-968 [FREE Full text] [doi: 10.1177/1545968315575612] [Medline: 25782693]
145. Alon G, Levitt AF, McCarthy PA. Functional electrical stimulation (FES) may modify the poor prognosis of stroke survivors with severe motor loss of the upper extremity: a preliminary study. Am J Phys Med Rehabil 2008 Aug;87(8):627-636. [doi: 10.1097/PHM.0b013e31817fabc1] [Medline: 18645322]
146. Yuzer GF, Dönmez B, Özgirgin N. A randomized controlled study: effectiveness of functional electrical stimulation on wrist and finger flexor spasticity in hemiplegia. J Stroke Cerebrovasc Dis 2017 Jul;26(7):1467-1471. [doi: 10.1016/j.jstrokecerebrovasdis.2017.03.011] [Medline: 28462794]
147. Villafañe JH, Taveggia G, Galeri S, Bissolotti L, Mullè C, Imperio G, et al. Efficacy of short-term robot-assisted rehabilitation in patients with hand paralysis after stroke: a randomized clinical trial. Hand (N Y) 2018 Jan;13(1):95-102 [FREE Full text] [doi: 10.1177/1558944717692096] [Medline: 28719996]
148. Friedman N, Chan V, Reinkensmeyer AN, Beroukhim A, Zambrano GJ, Bachman M, et al. Retraining and assessing hand movement after stroke using the MusicGlove: comparison with conventional hand therapy and isometric grip training. J Neuroeng Rehabil 2014 Apr 30;11:76 [FREE Full text] [doi: 10.1186/1743-0003-11-76] [Medline: 24885076]
149. Prange-Lasonder G, Radder B, Kottink AI, Melendez-Calderon A, Buurke J, Rietman J. Applying a soft-robotic glove as assistive device and training tool with games to support hand function after stroke: Preliminary results on feasibility and potential clinical impact. IEEE Int Conf Rehabil Robot 2017 Jul;2017:1401-1406. [doi: 10.1109/ICORR.2017.8009444] [Medline: 28814016]
150. Barry JG, Ross SA, Woehrle J. Therapy incorporating a dynamic wrist-hand orthosis versus manual assistance in chronic stroke: a pilot study. J Neurol Phys Ther 2012 Mar;36(1):17-24. [doi: 10.1097/NPT.0b013e318246203e] [Medline: 22354108]
151. Knutson JS, Gunzler DD, Wilson RD, Chae J. Contralaterally controlled functional electrical stimulation improves hand dexterity in chronic hemiparesis: a randomized trial. Stroke 2016 Oct;47(10):2596-2602 [FREE Full text] [doi: 10.1161/STROKEAHA.116.013791] [Medline: 27608819]
152. Cramer SC. Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. Ann Neurol 2008 Mar;63(3):272-287. [doi: 10.1002/ana.21393] [Medline: 18383072]
153. Church G, Parker J, Powell L, Mawson S. The effectiveness of group exercise for improving activity and participation in adult stroke survivors: a systematic review. Physiotherapy 2019 Jan 19. [doi: 10.1016/j.physio.2019.01.005] [Medline: 31003848]
154. Stroke Association. 2017. State of the Nation: Stroke Statistics. URL: https://www.stroke.org.uk/resources/state-nation-stroke-statistics [accessed 2019-11-07]
155. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002 Mar;14(2):109-114. [doi: 10.1097/00002281-200203000-00006] [Medline: 11845014]
156. Akobeng AK. Understanding randomised controlled trials. Arch Dis Child 2005 Aug;90(8):840-844 [FREE Full text] [doi: 10.1136/adc.2004.058222] [Medline: 16040885]
157. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. Soc Sci Med 2012 Dec;75(12):2299-2306. [doi: 10.1016/j.socscimed.2012.08.032] [Medline: 22989491]
158. Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA: Sage; 1997.
159. Moore G. Cramming more components onto integrated circuits. Proc IEEE 1998;86(1):82-85. [doi: 10.1109/jproc.1998.658762]
160. National Institute of Health and Care Excellence (NICE). 2019. 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 [accessed 2019-11-07]

Abbreviations

ARAT: Action Research Arm Test
BBT: Box and Blocks Test
BI: Barthel Index
CAHAI: Chedoke Arm and Hand Activity Inventory
CRoB: Cochrane Risk of Bias

DHTs: Digital Health Technologies
FES: functional electrical stimulation
ICF: International Classification of Functioning, Disability, and Health
ICT: information and communication technology
JTHFT: Jebsen-Taylor Hand Function Test
MAL: Motor Activity Log
MAS: Motor Assessment Scale
MCID: minimally clinically important difference
MeSH: Medical Subject Headings
NIHR: National Institute for Health Research
QuickDASH: short version of disabilities of arm, shoulder, and hand
RCT: randomized controlled trial
RMA: Rivermead Motor Assessment
SIS: Stroke Impact Scale
UEFT: Upper Extremity Function Test
UL: upper limb
WHO: World Health Organization
WMFT: Wolf Motor Function Test

Edited by G Eysenbach; submitted 03.09.19; peer-reviewed by I Cikajlo, E Sadeghi-Demneh, H Rostami, X Lu, F Lanfranchi, S Haghayegh; comments to author 28.09.19; revised version received 16.10.19; accepted 22.10.19; published 08.01.20.

Please cite as:
Parker J, Powell L, Mawson S
Effectiveness of Upper Limb Wearable Technology for Improving Activity and Participation in Adult Stroke Survivors: Systematic Review
J Med Internet Res 2020;22(1):e15981
URL: https://www.jmir.org/2020/1/e15981
doi:10.2196/15981
PMID:31913131

Original Paper

Valuable Genomes: Taxonomy and Archetypes of Business Models in Direct-to-Consumer Genetic Testing

Scott Thiebes[1*], MSc; Philipp A Toussaint[1*], BSc; Jaehyeon Ju[2], PhD; Jae-Hyeon Ahn[3], PhD; Kalle Lyytinen[4], PhD; Ali Sunyaev[1], PhD

[1]Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2]Desautels Faculty of Management, McGill University, Montreal, QC, Canada

[3]College of Business, Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea

[4]Department of Design and Innovation, Case Western Reserve University, Cleveland, OH, United States

[*]these authors contributed equally

Corresponding Author:
Ali Sunyaev, PhD
Department of Economics and Management
Karlsruhe Institute of Technology
Kaiserstrasse 89
Karlsruhe, 76133
Germany
Phone: 49 721608460
Email: sunyaev@kit.edu

Abstract

Background: Recent progress in genome data collection and analysis technologies has led to a surge of direct-to-consumer (DTC) genetic testing services. Owing to the clinical value and sensitivity of genomic data, as well as uncertainty and hearsay surrounding business practices of DTC genetic testing service providers, DTC genetic testing has faced significant criticism by researchers and practitioners. Research in this area has centered on ethical and legal implications of providing genetic tests directly to consumers, but we still lack a more profound understanding of how businesses in the DTC genetic testing markets work and provide value to different stakeholders.

Objective: The aim of this study was to address the lack of knowledge concerning business models of DTC genetic testing services by systematically identifying the salient properties of various DTC genetic testing service business models as well as discerning dominant business models in the market.

Methods: We employed a 3-phased research approach. In phase 1, we set up a database of 277 DTC genetic testing services. In phase 2, we drew on these data as well as conceptual models of DTC genetic testing services and iteratively developed a taxonomy of DTC genetic testing service business models. In phase 3, we used a 2-stage clustering method to cluster the 277 services that we identified during phase 1 and derived 6 dominant archetypes of DTC genetic testing service business models.

Results: The contributions of this research are 2-fold. First, we provided a first of its kind, systematically developed taxonomy of DTC genetic testing service business models consisting of 15 dimensions in 4 categories. Each dimension comprises 2 to 5 characteristics and captures relevant aspects of DTC genetic testing service business models. Second, we derived 6 archetypes of DTC genetic testing service business models named as follows: (1) low-cost DTC genomics for enthusiasts, (2) high-privacy DTC genomics for enthusiasts, (3) specific information tests, (4) simple health tests, (5) basic low-value DTC genomics, and (6) comprehensive tests and low data processing.

Conclusions: Our analysis paints a much more complex business landscape in the DTC genetic testing market than previously anticipated. This calls for further research on business models and their effects that underlie DTC genetic testing services and invites specific regulatory interventions to protect consumers and level the playing field.

(J Med Internet Res 2020;22(1):e14890) doi:10.2196/14890

KEYWORDS

genomics; genetic testing; genetic privacy; direct-to-consumer screening and testing; taxonomy; cluster analysis

Introduction

Background and Objectives

When 23andMe opened its Web store to the public in 2007, it was among the first in the second wave of the so-called direct-to-consumer (DTC) genetic testing services that publicized the idea of personal genomics [1]. The surge of DTC genetic testing services such as 23andMe, Ancestry.com, or FamilyTreeDNA, to name a few, during the last decade is thereby mainly driven by the rapidly declining costs for collecting and analyzing genome data [2] and the public's growing interest in genomics [3]. Overall, estimates suggest that the global DTC genetic testing market will be worth approximately US $610 million by 2026 [4].

DTC genetic testing refers to genetic tests targeted toward consumers that do not require the involvement of a medical professional in mediating the service and interpreting it [5]. Typical tests offered by DTC genetic testing services include ancestry tests, nonmedical lifestyle tests (eg, traits, fitness, nutrition), and medical tests (eg, carrier status, genetic health) [3]. Compared with conventional medical testing, DTC genetic tests are, however, initiated by consumers, not their physicians, and sold directly to consumers over the internet [6]. Consequently, consumers are responsible for controlling and managing their genetic information on the Web, as well as choosing an interpreter for their genetic information or interpreting it themselves [6]. Despite the ongoing proliferation of genetic testing and the public's rising interest therein, DTC genetic testing has been a source of controversy ever since the inauguration of the first services offering such tests [5]. This especially pertains to business practices underlying DTC genetic testing services, as it is often believed that such services resell access to their consumers' genetic data to increase revenue and to compensate for selling budget-priced tests. 23andMe, for example, uses its DTC front end to build a health information database and sells access to this database to clinical research and biopharmaceutical companies [7]. Concerns about the impact and business practices of DTC genetic testing services climaxed in several professional associations recommending to consumers to refrain from using DTC genetic testing services at all [5] and the US Food and Drug Administration (FDA) sending out cease and desist letters to several DTC genetic testing services in 2013 [6].

To realize the full potential of a disruptive technology such as genetic testing, it needs to be met by innovative business models [8]. Toward this end, DTC genetic testing can be regarded as a class of business models that can serve to make genetic testing more affordable and accessible. In light of the glaring contradiction between the ongoing proliferation of DTC genetic testing services, on the one hand, and the spreading skepticism toward their business practices on the other hand [9], however, there still seems to be little consensus among providers of such services, policy makers, professionals, and the general public on what actually constitutes suitable business models for DTC genetic testing. Several DTC genetic testing services, for example, recently started to shift their business models toward greater involvement of medical professionals and test approvals by regulatory bodies [6]. This resulted in the FDA-granting approvals for certain DTC genetic tests [6], which is likely to provide additional impetus to this market. Thus far, research provides little guidance for assessing or developing business models in the context of DTC genetic testing. Owing to the clinical value and sensitivity of genome data, past research on DTC genetic testing has primarily focused on ethical and legal issues surrounding genetic tests offered directly to consumers [10], including impact and clinical utility of DTC genetic tests [11], as well as studies of awareness and perceptions of DTC genetics and its risks [12-14]. Most of our knowledge about business models of DTC genetic testing services today stems from newspaper articles [15], blogposts discussing specific DTC genetic testing services [7], or company reports [16]. To the best of our knowledge, so far only a single white paper explicitly investigates business models underlying DTC genetic testing services (cf, the Business Models in Genomics section), and there are no published academic articles on this topic. A first step toward closing this knowledge gap pertains to scrutinizing the status quo and understanding what characterizes business models in the DTC genetic testing context today as well as what, if any, dominant business models have emerged in this market. We therefore ask the following research questions:

- RQ1: What are salient properties of DTC genetic testing service business models?
- RQ2: What dominant business models of DTC genetic testing services have emerged?

To answer these research questions, we develop a taxonomy of DTC genetic testing service business models (RQ1) and subsequently draw on this taxonomy to derive archetypes of DTC genetic testing service business models (RQ2). In doing so, we provide a systematic classification of DTC genetic testing service business models and expose their essential characteristics. We posit that the knowledge encapsulated in the taxonomy and dominant archetypes makes fruitful contributions to research on the role and impact of building innovative (internet-based) business models in health care. It can aid policy makers to better understand how DTC genetic testing services operate and design policies and regulations to protect consumers and enable fair competition when possible while realizing their full potential. Moreover, it should also help addressing health professionals' concerns regarding the impact of these services and increase consumers' awareness to make more informed decisions about using such services. Finally, for nascent businesses in the DTC genetic testing industry, our findings can serve as a blueprint of the emerging competitive landscape and as a roadmap to advance their business models.

Business Models in Genomics

Despite the term business model being commonly used in management and strategy, little consensus prevails over what constitutes a business model [17]. Extant literature has proposed an abundance of definitions for business models [8,18,19]. Drawing on Shafer et al [17], who synthesized extant conceptualizations of business models within scientific literature, we understand a business model as "a representation of a firm's underlying core logic and strategic choices for creating and capturing value within a value network." Consequently, business

models comprise 4 major categories of components: (1) *strategic choices* (eg, customers, target markets, value propositions, revenues and pricing, competitors), (2) *value creation* (eg, key resources, assets, processes), (3) *value network* (eg, information and product flows between an organization, its suppliers, and customers), and (4) *capturing value* (eg, profit-making mechanisms).

As pointed out earlier, disruptive technologies like genetic testing must be met by innovative business models to realize their full potential [8]. However, the health care sector is knowingly slow in adopting disruptive technologies, especially the innovative business models that these technologies afford [8,20]. Looking at genetic testing, we see that this nascent but arguably increasingly important aspect of modern health care is no exception to this phenomenon. Toward this end, scientific literature on the business aspects of DTC genetic testing services, especially their underlying business models, remains scarce [21]. The stream of research that is closest to this topic relates to socioeconomic research on DTC genetic testing [1]. Research in this stream is concerned with the marketing strategies of DTC genetic testing services and their impact on consumers [22,23], as well as economic implications of consumers freely sharing their genome data [24-26]. Extant literature has also covered the history [27] and size [28] of the global DTC genetic testing market. Our review of the related literature resulted, however, in only 1 white paper by Vanhala and Reijonsaari [29] that explicitly investigates business models in the context of DTC genetic testing. They define a business model as connecting consumers' needs with the solution offered by a service. Accordingly, any DTC genetic testing service business model comprises value propositions, distribution channels, revenue logic, customer segments, and key resources (eg, genome data). On the basis of these 5 business model aspects, Vanhala and Reijonsaari [29] further derive 5 categories of DTC genetic testing service business models: (1) comprehensive genomic tests for consumers and as genome data bank material (eg, 23andMe), (2) genomics as part of individual health planning (eg, MD Revolution), (3) genomic services based on comprehensive genetic testing (eg, Genetrainer), (4) medical precision tests for consumers (eg, Myriad Genetics), and (5) restricted trait tests (eg, Genecodebook Oy) [29]. Although this categorization certainly is a valuable step toward shedding light onto DTC genetic testing service business models, the authors provide little information on the employed methodology and base their analysis on assessing only a limited number of DTC genetic testing services. Adding to this, the study was published in 2013 and thus does not account for recent changes in the landscape of DTC genetic testing service business models.

Methods

Overview

To answer our research questions, we adopted the 3-phased approach of Remane et al [30]. A detailed description of each phase is given later, and Table 1 provides a brief summary of the individual phases.

Table 1. Overview of the research approach.

Phases	Phase 1: Database setup	Phase 2: Taxonomy development	Phase 3: Cluster analysis
Inputs	• Desk research • Web-based genetic testing service repositories [31-33]	• DTC[a] genomics literature (deductive iterations) • List of DTC genetic testing services (inductive iterations)	• DTC genetic testing service business model taxonomy
Steps	• Compile a list of DTC genetic testing services • Filter services not available anymore • Collect information about services from multiple sources (eg, websites, blogs, research, or news articles)	• Define a meta-characteristic • Develop taxonomy iteratively until all ending conditions are met	• Identify suitable numbers of clusters (Ward's method) • Run iterative partitioning algorithm with the identified numbers of suitable clusters • Select the most fit cluster solution • Analyze cluster solution and derive archetypes
Outcomes	• List of 277 DTC genetic testing services	• Taxonomy of DTC genetic testing services' business models with 41 characteristics in 15 dimensions	• 6 archetypes of DTC genetic testing service business models

[a]DTC: direct-to-consumer.

Phase 1: Database Setup

In phase 1, a database of DTC genetic testing services was set up. It served as a basis for taxonomy development and cluster analysis in phases 2 and 3, respectively. As we are solely interested in DTC genetic testing services that have a Web presence, the internet provided a good starting point to collect information about these kinds of services. We therefore conducted a desk search using the internet to set up our database. Our search led to the identification of 3 central resources that list a large body of genomic service providers (not limited to DTC genetic testing services). First, Phillips [31] has compiled a comprehensive list of 301 DTC genetic testing services, which is available on her website. Second, DNA Testing Choice [32], a UK-based news and reviews service of genetic tests, offers rankings for a variety of DTC genetic tests. As some DTC genetic testing services may offer more than one test, they may also appear in multiple rankings on DNA Testing Choice.

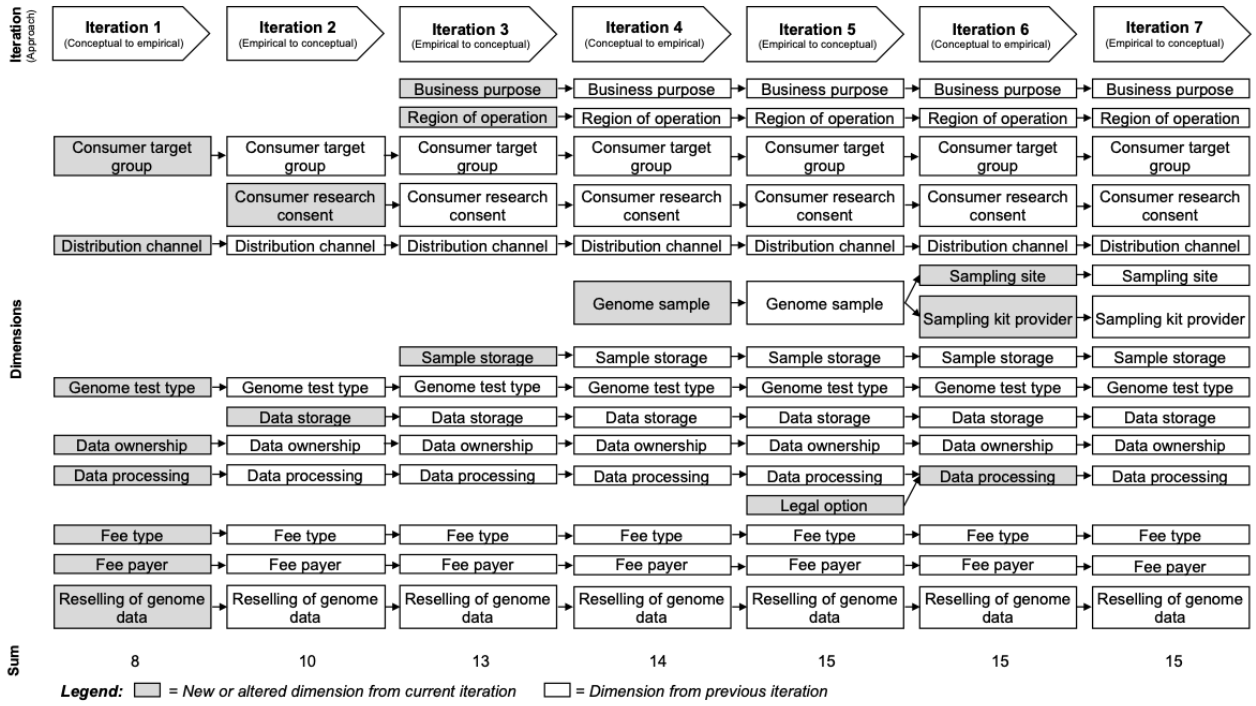
Finally, the International Society of Genetic Genealogy [33] also provides a list of DTC genetic testing services. Overall, all 3 sources together listed 428 genetic testing services. For each service, we noted its name, website, a brief service description, and relevant sources (ie, websites that reference the respective service). As the focus of this study is on DTC genetic testing services, we screened all entries for DTC genetic tests before the taxonomy development process. This led to the exclusion of 171 services that were either not available anymore or did not fit our definition of DTC genetic testing services (eg, services that offered genetic testing of animals). We further identified 20 additional DTC genetic testing services while collecting information about services already included. In total, our final database contains 277 DTC genetic testing services. A complete list of services included in and excluded from our database can be found in Multimedia Appendix 1.

Phase 2: Taxonomy Development

The second phase focused on the development of a taxonomy of DTC genetic testing service business models (thus answering RQ1). Taxonomies are important tools in scientific disciplines as they provide researchers with fundamental categories to analyze and understand complex domains [34]. We based the development of our taxonomy on the method by Nickerson et al [34]. It provides a systematic approach (as opposed to an ad hoc approach) to taxonomy development that combines inductive and deductive reasoning and has been extensively used to systematically develop taxonomies for phenomena in health care [35-42] and other domains [30,34,43,44]. The method consists of 7 iterative steps and provides guidelines for each step in the taxonomy development process. Step 1 abides the selection of a so-called meta-characteristic. Being the most comprehensive characteristic, the meta-characteristic serves to avoid a situation of naïve empiricism, acting as the basis for the choice of characteristics included in the taxonomy [34]. Each characteristic in the taxonomy must therefore be a logical consequence of the meta-characteristic [34]. Drawing on RQ1, we defined our meta-characteristic as *salient properties of business models of DTC genetic testing services*. As steps 3 to 7 of the taxonomy development are iterative, some predetermined conditions that end the process must be defined in step 2. For this research, we adopted the 5 subjective ending conditions (ie, conciseness, robustness, comprehensiveness, extendibility, and explanatory) and 8 objective ending conditions (eg, characteristics in a dimension are mutually exclusive, and there were no changes made to the taxonomy during the last iteration) provided by Nickerson et al [34]. Next, steps 3 to 7 involve the iterative development of the taxonomy (cf, Multimedia Appendix 2), whereby each iteration starts by choosing either a conceptual-to-empirical (deductive) or empirical-to-conceptual (inductive) approach in step 3 and ends with a review of whether the ending conditions are met in step 7 [34]. If the ending conditions are not met, an additional iteration is performed.

In total, we performed 7 iterations. An overview of the development of dimensions for individual iterations is given in Figure 1. The first iteration followed the conceptual-to-empirical approach and was based on the model of Vanhala and Reijonsaari [29]. It led to an initial taxonomy with 8 dimensions. We classified DTC genetic testing services examined by Vanhala and Reijonsaari [29] and the top 10 ancestry, paternity, and health test service providers according to DNA Testing Choice [32] in the second iteration to verify the validity of our initial taxonomy. This led to the addition of 2 new dimensions.

Figure 1. Overview of the taxonomy development iterations.



Before the next iteration, we grouped all services based on the tests they offered. Similar to DNA Testing Choice's ranking categories, we grouped services into 1 of the 3 categories: (1) *genomics enthusiasts* (eg, ancestry, genetic dating, trait testing), (2) *relationship tests* (eg, paternity, maternity), and (3) *health tests*. Like iteration 2, iteration 3 also followed the empirical-to-conceptual approach. After the analysis of the *genomics enthusiasts* subset, we added 3 dimensions. At this point, a review of dimensions and characteristics in iteration 4 (conceptual-to-empirical) resulted in the addition of 1 further dimension. In iteration 5 (empirical-to-conceptual), we analyzed all services offering *relationship tests*, which led to the addition of 1 new dimension. The following iteration 6 again followed the conceptual-to-empirical approach. Within this iteration, we merged 2 dimensions and split 1 dimension into 2 distinct dimensions. During iteration 7, the remaining subset of *Health Tests* was examined, which led to no alterations to the taxonomy. At this stage, the taxonomy fulfilled all ending conditions after 7 iterations (cf, Multimedia Appendix 3).

Phase 3: Cluster Analysis

The third phase focused on deriving archetypes of DTC genetic testing service business models (RQ2) using the previously created taxonomy as a baseline for cluster analysis. Cluster analysis is a process of finding distinct groups of objects (ie, clusters) in data [45]. The objective is to find groups (clusters) for which the objects of 1 group are highly similar in selected attributes, whereas they are as dissimilar as possible from objects in the other groups [45]. An abundance of clustering methods is available, and choosing the most suitable method can be cumbersome and error prone in terms of what similarity or dissimilarity measure to choose, how many clusters to generate, or overall performance of clustering algorithms [46]. Iterative partitioning algorithms such as *k*-means, for example, provide better performance than hierarchical clustering methods, but in turn usually require the a priori definition of how many clusters to produce [46]. To address these problems, we adopted the 2-stage clustering approach suggested by Punj and Stewart [46]. In this process, the first stage utilizes a hierarchical method to determine a preliminary solution, which can be used to deduce candidate numbers of clusters that serve as starting points for the iterative partitioning algorithm in stage 2. On the basis of the candidate number of clusters obtained in stage 1, an iterative partitioning algorithm arranges the included objects in their final cluster solution in stage 2 [46]. According to Remane et al [30], we utilized Ward's method for stage 1 and the *k*-means algorithm for stage 2.

Ward's method is an agglomerative clustering procedure. It starts by combining the 2 objects closest to each other into 1 cluster and repeats this process until all objects belong to the same cluster [47]. During each iteration, the similarity between 2 clusters is calculated by the number of identical characteristics. As our taxonomy holds only binary data (ie, a characteristic is either applicable or not), the squared Euclidean distance, which places progressively greater weight on objects that are further apart from each other, is a suitable similarity measure [30]. The dendrogram produced by Ward's method indicated that a 4-, 5-, or 6-cluster solution would have the most explanatory power for our dataset. Also, by reviewing the scree plot, the elbow rule suggested the 4-, 5-, 7-, or 9-cluster solution in this particular order. With the preliminary cluster solutions in place, we used the *k*-means method to derive our final cluster solution in the second stage. The *k*-means method produces a partition of the dataset into an a priori defined number of clusters [47]. Starting with an initial partition, objects are moved into other clusters if they are closer to its mean vector than that of their current cluster, usually calculated using Euclidean distance. After each iteration, the mean vectors are updated [47]. The procedure continues until all objects are closer to the mean vector of their own cluster than to the mean vectors of any of the other clusters or no significant changes are found [47]. For the 4, 5, 6, 7, and 9 cluster solutions, the algorithm ran through 10, 17, 9, 13, and 12 iterations before achieving convergence, respectively. However, retrieved significance values for characteristics (ie, how relevant a certain characteristic is for the cluster solution) indicated that the 4 and 5 cluster solutions were of inadequate quality because they possessed too many irrelevant characteristics (cf, Multimedia Appendix 4). We therefore did not consider those cluster solutions any further. The remaining cluster solutions (ie, 6, 7, and 9 clusters) were next manually compared for their explanatory power by 2 researchers (ie, we sought to find meaningful interpretations for all clusters in all cluster solutions). Although we were able to find meaningful interpretations for the 7-cluster solution, it produced entirely different clusters, which we deemed less meaningful and providing little to no additional insight compared with the 6-cluster solution. For the 9-cluster solution, on the other hand, we were unable to find meaningful interpretations for all clusters. Hence, we selected the 6-cluster solution as the most suitable one for this study and report it below.

Results

Direct-to-Consumer Genetic Testing Service Business Models Taxonomy

Our final taxonomy consists of 15 dimensions. Each dimension consists of 2 to 4 characteristics, with a total of 41 characteristics. Furthermore, the dimensions have been grouped into 4 categories based on the previously outlined categories of components of business models (ie, strategic choices, value network, create value, and capture value) to provide a better understanding of how the dimensions relate to each other. Table 2 offers an overview of the final taxonomy, whereas the results of our coding can be found in Multimedia Appendix 5. In the following, we describe each dimension and its characteristics in detail.

Table 2. Taxonomy of direct-to-consumer genetic testing services' business models.

Dimension	Characteristics
Strategic choices	
Business purpose	For profit; nonprofit
Region of operation	Local; worldwide
Consumer target group	Enthusiasts; specific information seekers; enthusiasts and specific information seekers; chronic health issue and risk group
Consumer research consent	Mandatory; optional; data not used
Value network	
Distribution channel	Internet only; health care professionals only; multicontact service
Sampling site	Home collection; lab collection; home and lab collection
Sampling kit provider	Service provider; third party; service provider and third party
Sample storage	Never; mandatory; consumer decision
Create value	
Genome test type	Genotyping; sequencing; genotyping and sequencing
Data storage	No storage; isolated storage; database for service provider
Data ownership	Consumer; service provider
Data processing	No interpretation; basic interpretation; value-added interpretation
Capture value	
Fee type	Pay-per-use; pay-per-use and subscription; no fee
Fee payer	Consumer only; consumer and health insurance
Reselling of genome data	Yes; no

Strategic Choices

The first category of components of a business model entails 4 dimensions related to strategic choices of a DTC genetic testing service, such as target markets and customers. The *business purpose* dimension answers the basic question of whether a service provider seeks to generate profit or whether it is a nonprofit organization that contributes to genomics research and makes DTC genetic testing more accessible to consumers. The *region of operation* dimension answers the question where a DTC genetic testing service is offered. In our taxonomy, we distinguish between the characteristics local (ie, tests are only available in the country the service has been registered in) and worldwide (ie, tests are offered all over the world with the exception of countries that do not allow such tests by law). The third dimension, *consumer target group*, divides DTC genetic testing services into 4 target groups. Products aimed at enthusiasts seek to spark the curiosity of the consumer for information on their DNA such as ancestry, taste tests, lactose intolerance, or health traits [29]. The second target group concerns specific information seekers. Services for this target group offer genetic tests that aim to answer a specific question and include, for example, paternity and other relationship tests, immigration tests, or tests for certain diseases. There are also genetic testing services that target both enthusiasts and specific information seekers and are summarized in the third characteristic of this dimension. The last target group addresses consumers dealing with chronic health issues and related risks (eg, individuals with diabetes, individuals with high blood pressure, or individuals with an increased risk of cancer). Finally, the *consumer research consent option* dimension indicates whether DTC genetic testing service providers ask for consumers' consent to use their data for research. For some services, it is mandatory that consumers give their consent for using their data for research to purchase the service. Other service providers either give consumers the option to consent into their personal genome data being used for research or do not utilize their customers' data for research purposes at all.

Value Network

The value network category comprises 4 dimensions that characterize a DTC genetic testing service's relationship to key partners and customers, as well as how information, products, and services flow through this network. The *distribution channel* dimension describes how products and services are communicated and offered to the consumer. Although many DTC genetic testing service providers only offer an internet presence for all consumer means, some service providers may require a health care professional to be involved. However, health care professionals mainly serve as a means for distributing and, in some cases, carrying out sample collection and follow-up patient counseling but are responsible neither for performing the actual genetic testing nor for interpreting test results. The third and most comprehensive distribution channel is described as a multicontact service [29]. This includes internet solutions, mobile apps, telephone consulting, stores, and home visits to offer the product to the consumers. The s*ampling site* dimension summarizes where a consumer's genome sample is collected.

Many services offer the collection of genetic material via home collection kits that are mailed to their consumers. The sample is then taken by the consumers themselves (eg, buccal swab or saliva sample) and sent back for analysis. Other services require their consumers to visit a lab, where samples are taken by the staff. Lab collection is usually offered either as a convenience to the consumer or because it is legally required (eg, a paternity test that is to be acknowledged by the court). Some services offer home and lab collection of genome samples. Adding to this, the *sampling kit provider* dimension describes whether a service provider offers their own sample collection kit or whether a third-party sample collection kit is used. Some services offer both, their own sample collection kit and the option to use a third-party kit. Finally, the *sample storage* dimension indicates whether the collected sample is destroyed (ie, never stored) or kept after the genome data have been generated. Although for some genetic testing services it is mandatory that genetic samples are kept (eg, for legally binding paternity tests), others leave this decision up to their customers.

Create Value

The third category contains 4 dimensions, which describe how DTC genetic testing services create value for customers. It thus focuses on their products and processes. The *genome test type* dimension determines what kind of method is used to generate genome data and consists of three characteristics. A service provider may either offer genotyping only, sequencing (whole genome or whole exome) only, or both, genotyping and sequencing. Next, the *data storage* dimension considers how DTC genetic testing services store the genome data of their consumers. Service providers can either not store the produced data, store it isolated for the consumers' access only, or store it in a common database, which is used to improve service quality. If the data are not stored, then they are deleted shortly after the consumers retrieve their genome data. Some services collect fees for the isolated storage of the data and will only keep it as long as the consumers decide to store it. The *data ownership* dimension classifies DTC genetic testing services in terms of who owns the genome data. By purchasing a product from a service, the consumers agree to their terms of service. The terms of service usually state whether the collected genome data are the property of the consumers or the service provider. If the ownership stays with the consumers, the decision power over the data also remains with the consumers. If a business claims ownership of the data, then it is authorized to use the data without restrictions to further the company's interests. Finally, the fourth and last dimension, *data processing*, describes the degree of genome data analysis provided by a service. Some DTC genetic testing services do not offer any interpretation but instead deliver the raw genome data by means of genotyping or DNA sequencing, only. If services offer interpretation of the produced genome data, they usually create reports on certain information. Most services provide an analysis of the genome data in terms of ancestry information, health traits, paternity tests, or cancer tests. Alternatively, some services offer value-added interpretation, whereby they augment their interpretation with additional services such as, for example, carrying out legally binding paternity tests (as opposed to cheaper nonlegally binding tests) or providing diet plans or supplements based on the analysis of their consumers' genome.

Capture Value

Capture value is the final category of business model components. It pertains to three dimensions that describe how DTC genetic testing service providers generate revenue. The *fee-type* dimension is concerned with the providers' pricing model. Consumers may be charged on a per-use basis or on a per-use basis paired with a subscription model (eg, to account for additional services besides the actual test). Some tests are offered to consumers free of charge. The *fee-payer* dimension describes who pays for the offered tests. Tests can be paid for entirely by consumers themselves as well as partly or completely by the consumers' health insurances. The *reselling of genome data* dimension relates to whether a service provider generates revenue from reselling collected genome data to third parties. Possible customers for genome data include research institutes, clinics, or pharmaceutical companies.

Archetypes of Direct-to-Consumer Genetic Testing Services' Business Models

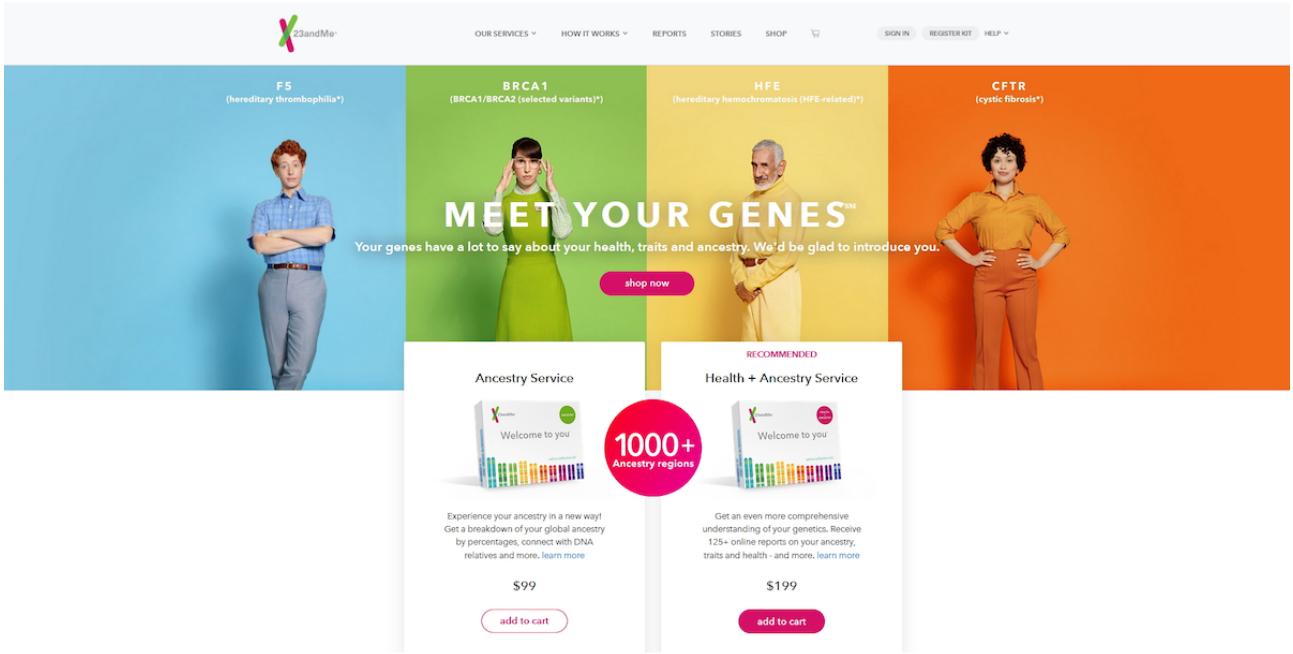
The clusters in the 6-cluster solution comprise between 21 and 73 of the 277 DTC genomics services in our database. Thereby, each cluster has a different focus regarding the dimensions and characteristics of the DTC genetic testing service business model taxonomy. As the taxonomy development method by Nickerson et al [34] results in characteristics that are mutually exclusive and collectively exhaustive, the data can be interpreted as percentages. For example, 89% (42/47) of the companies in cluster 1 operate worldwide whereas 11% (5/47) offer services in their respective country only. Table MA6-1 in Multimedia Appendix 6 provides an overview of the results of the cluster analysis, whereby darker colors represent higher percentages of services in the cluster belonging to a characteristic for the corresponding dimension. We elaborate on each cluster below by highlighting its most representative characteristics and providing examples of typical DTC genomics services.

Cluster 1: Low-Cost Direct-to-Consumer Genomics for Enthusiasts

The first cluster combines the low-cost genotyping tests offered by companies like 23andMe (see Figure 2 for a screenshot of their website) with the costlier sequencing tests provided by companies like Veritas Genetics who strive to make sequencing affordable for the enthusiast DTC market. These companies operate worldwide as they provide their services over the internet and less via multicontact channels. Most operators like 23andMe only charge the consumers a 1-time fee for their home collection kit but claim the rights of the produced genome data and utilize it to improve their own services and resell it for profit to compensate for the relatively low prices of their tests. Other services like Ancestry.com operate on a pay-per-use and subscription fee model and therefore leave the ownership of the data with the consumer. Finally, some services offer health-related tests that are eligible for insurance coverage. Although the offered services are value-added tests, services either make it mandatory or give customers the option to willingly participate in research with their data. Other prominent

examples of services in this cluster are FamilyTreeDNA, MyHeritageDNA, or the Genographic Project.

Figure 2. Screenshot of the 23andMe website from October 2019.



Cluster 2: High-Privacy Direct-to-Consumer Genomics for Enthusiasts

Cluster 2 is in many aspects similar to cluster 1 but differentiates itself in the crucial dimensions concerning the customers' data privacy. Most services in cluster 2, such as African Ancestry, provide internet only or internet and telephone solutions for genomic enthusiasts. African Ancestry provides a simple genotyping test that answers the general question of whether a customer has origins tracing to Africa and other information regarding their African ancestry. The tests can only be taken with a home collection kit and results are made available on the Web. The generated genome data remain the property of the consumer for all services in this cluster, and the genome data are either erased or stored in an isolated storage accessible by the consumer only. It is not sold for revenue. Furthermore, the data are usually not used for research, but consumers may be given the option to provide their data. The costs for these tests are covered by a 1-time fee for the consumer or an additional subscription. Other examples of services in this cluster include EasyDNA, FitGenes, or The Makings of Me.

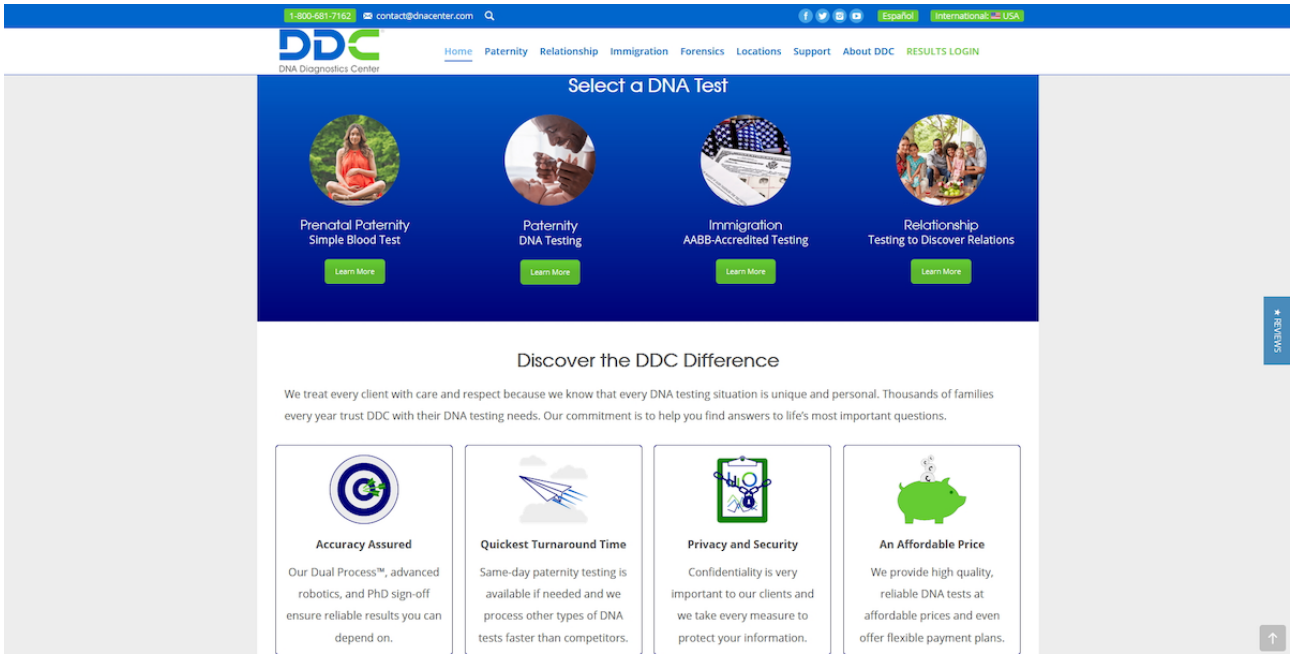
Cluster 3: Specific Information Tests

The third cluster holds DTC genetic testing services that provide consumers with solutions to specific questions and aim to provide a more elaborate service with value-added data processing. One prominent example for this group is DNA Diagnostics Center (see Figure 3 for a screenshot of their website). Their choice of distribution channel over the internet and via telephone allows for a worldwide service. DNA Diagnostics Center offers a variety of genome tests on specific information ranging from relationship to forensics on a 1-time cost basis for the customer. Although all genome tests of this cluster cover genotyping, the sample collection can either be performed by the consumers themselves with a service-provided home collection kit or by a professional (eg, for a legal paternity test at a laboratory operated by DNA Diagnostics Center). The genome data are primarily stored in isolation and sometimes used to improve service quality (ie, the data are stored in a database for later reference during the genotyping process), whereas the consumers' genome data are not used for inhouse research or sold to a third party for revenue. Moreover, it is often mandatory to store the sample (eg, for later reference of legal relationship tests). This cluster holds several representative service providers such as Alpha Biolabs, Dadchecksilver, or Who'z the daddy.

Cluster 4: Simple Health Tests

Cluster 4 combines chronic health-related genomic services such as Fulgent Diagnostics with the more casual health and wellness focused services such as SkinDNA Canada. These health-related services are either available through health professionals only or multicontact services, which may also include health professionals. This elaborate customer relationship leads to many services only operating locally, though some have a worldwide network. It comes as no surprise that this cluster contains most services aimed at chronic health issue and risk group consumers. Nevertheless, the cluster is still mainly populated with specific information seekers and enthusiast target groups. The 1-time fees are usually covered by the consumers but may be covered by insurances as well. The companies provide their own sampling kit and after genotyping, the sample is destroyed if storing it is not necessary. These services offer a lab collection option but come also with a home collection kit. Although the data ownership remains with the consumers and genome data are not sold for revenue or utilized in research, the results are mostly basic reports, which may require further interpretation by a local physician. Other examples of services in this cluster include International Bioscience or Pillcheck.

Figure 3. Screenshot of the DNA Diagnostics Center website from October 2019.



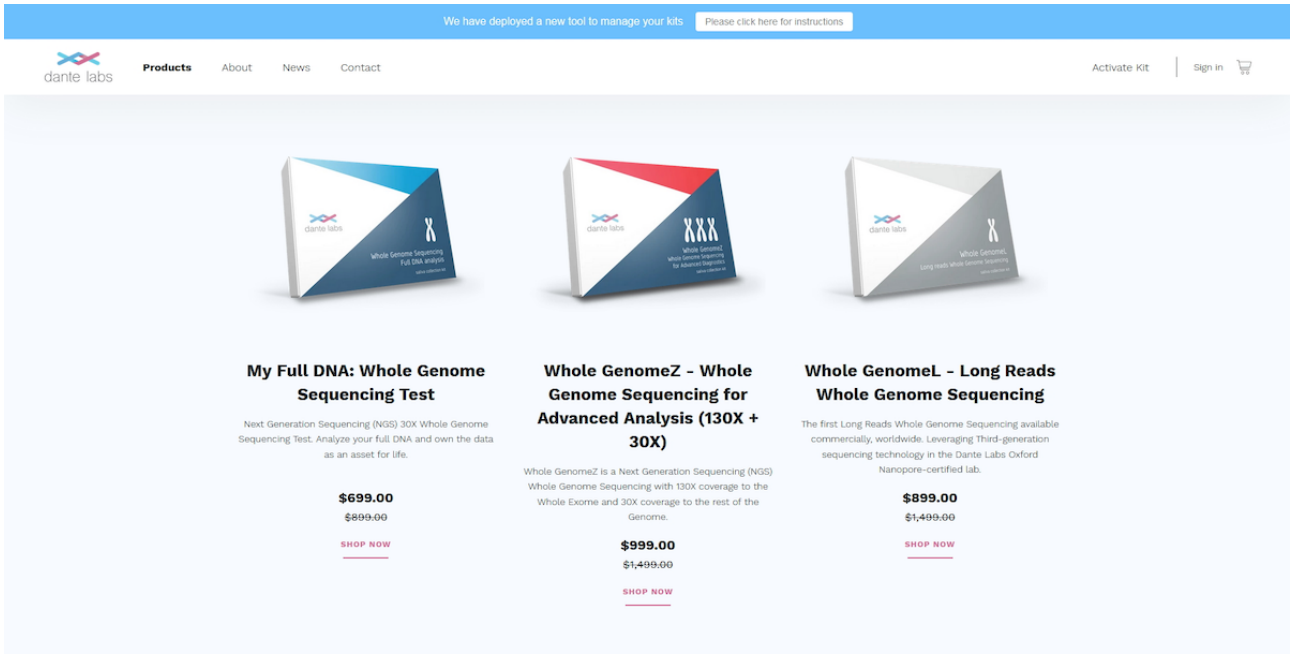
Cluster 5: Basic Low-Value Direct-to-Consumer Genomics

Cluster 5 differentiates itself from all other clusters in that it contains for profit as well as nonprofit companies that either charge a 1-time fee or offer services that are completely free. As services like Genetic Genie rely mostly on third-party sample collection, they can offer their services worldwide, exclusively via the internet as the consumers choose the sampling provider (eg, 23andMe) independently. The compatible sampling kits are for home collection only and aimed at genotyping tests. Consequently, the data ownership always lays with the consumers for services in this cluster and if stored, the genomic data are accessible for the costumers only. Furthermore, Genetic Genie does not sell the data for revenue, and there is no research consent option. As a result, the data processing is mostly done automatically, and the resulting interpretation is only of a basic nature. Nonetheless, enthusiasts may find a convenient way for additional insight into their genome. Other members of this cluster are Promethease, Roots for Real, or My Genetic Health.

Cluster 6: Comprehensive Tests and Low Data Processing

Offering sequencing only or sequencing and genotyping tests, the last cluster holds services focusing on the sample processing mostly (ie, offering no or basic interpretation of the genome data only). Dante labs (see Figure 4 for a screenshot of their website), for example, offers a worldwide whole genome sequencing home collection kit via their website with a basic health analysis and Web access to the consumers' entire sequenced genome. Other services in this cluster offer other forms of customer contact such as local labs or health professionals. The sampling kit may be provided by the service provider or additionally by a third party. Although this cluster has no specific target group, services are always paid for by the consumers with a 1-time fee. To compensate for the comprehensive genome tests and generate profit, companies often resell access to the produced genome data, make participation in research mandatory, or claim the right to use the data for company services. This is also mirrored in the data ownership, which may lie with the customer as well as with the company. Additional examples for services in this cluster are Full Genomes Corporation, Helix, or Genes for Good.

Figure 4. Screenshot of the Dante Labs website from October 2019.



Discussion

Principal Findings

Analysis of the taxonomy and derived archetypes unveils interesting insights into the current state of the DTC genetic testing market.

First, our results paint a much more heterogeneous landscape of the DTC genetic testing market than most of the extant literature in this area has conceived. This is not only highlighted because of the presence of 6 diverse business model archetypes that we identified but also supported by the fact that the archetypes' corresponding clusters are relatively evenly sized with cluster 5 being the smallest (n=21) and cluster 3 being the largest (n=73). Looking at the 5 business models described by Vanhala and Reijonsaari [29], the business model archetypes described here are rooted in a more diverse set of distinguishing dimensions (ie, 15 dimensions as opposed to 5 dimensions). This resulted in less (albeit still prevalent) emphasis on the different value propositions and consumer target groups and instead also included aspects such as major cost drivers (eg, dimensions such as sampling site and sampling kit provider). Furthermore, comparing the presented taxonomy with business model taxonomies of other disruptive technologies [30,48,49], we see that they share some similarities but also exhibit distinctive differences. Accordingly, our taxonomy includes several dimensions that are inherent to any business model, especially those dimensions that are related to customer segments, key partners, value propositions, or service pricing. Our fee-type dimension, for example, is comparable with the price structure dimension found in the carsharing business model taxonomy of Remane et al [30], albeit with slightly different characteristics. At the same time, however, the taxonomy presented here also includes several dimensions whose characteristics are more tailored toward the DTC genetic testing market (eg, distribution channel, fee payer), as well as dimensions that are entirely unique to the DTC genetic testing context (eg, genome test type).

Second, much of the controversy surrounding DTC genetic testing originates from concerns over the clinical value of such tests [50], consumers' capabilities of dealing with potentially misleading test results [11], and the assumption that DTC genetic testing services sell access to their consumers' genomic data to third parties [51]. To this end, cluster 1 seems to represent a business model archetype that many skeptics of DTC genetic testing services have in their minds when thinking of DTC genetic testing. This is further supported by the fact that the most prominent and probably most often criticized players in the DTC genetic testing market such as 23andMe, AncestryDNA, and FamilyTreeDNA can be found in this cluster. The business model archetype represented by cluster 1 also closely resembles the comprehensive genomic tests for consumers and as genome data bank material business model described by Vanhala and Reijonsaari [29]. Overall, however, our analysis showed that the majority of DTC genetic testing services do, for example, not resell access to their consumers' genomic data to third parties for revenue (237 of 277). Even more so, cluster 2 represents a business model archetype where special emphasis is placed on consumers' privacy. Several explanations might exist for these surprising, yet interesting, findings. Genomics, particularly DTC genetic testing, is still a relatively young business [28]. Next to the dimensions in the capture value category, the consumer research consent and data ownership dimensions are especially deeply related to a DTC genomics service provider's profit as both dimensions exert a strong influence on what service providers can and cannot do with their primary resource, the produced genomic data. In this regard, it is important to note that scale benefits are mostly on the DTC genetic testing services' side rather than the consumers' side, as service providers can use already produced data to improve their service quality. Some service providers might also seek to incentivize interested individuals to use their service to establish a large enough database of genomic data that they

can then use to develop complementary revenue streams, which do not directly involve selling access to genomic data to third parties (eg, use collected data to develop new drugs). Accordingly, some services might still be in a phase where growth is considered more important than short-term profit by the services' stakeholders. Finally, some services might anticipate further declining costs for genome data collection and analysis because of technological advances, eventually making them more profitable and changing the relative benefits of different business models in the long run. An example for such a service is Veritas Genetics, who seek to provide whole genome sequencing services costing less than US $1000 to their consumers.

The third interesting finding pertains to recent debates about shifts in DTC genetic testing services' strategies and the emergence of what some call DTC genomics 2.0. Compared with the currently prevailing DTC genomics paradigm, DTC genomics 2.0 is characterized by a greater involvement of regulatory bodies and health care professionals, a stronger separation between health and nonhealth tests, and improved support and counseling for consumers [6]. To this end, our clusters show a clear separation between services mainly targeting specific information seekers and people with chronic diseases (clusters 3 and 4), and those primarily targeting enthusiasts (clusters 1, 2, 5, and 6). Specifically, cluster 4 exhibits the largest number of services whose primary distribution channel are health care professionals and services whose tests are often paid for by consumers' insurances. Cluster 4 represents a business model archetype that can serve as a prime example for a shift toward DTC genomics 2.0. Although such services arguably blur the lines between what is traditionally considered DTC genetic testing (ie, genetic tests, directly sold to consumers via the internet) and other forms of genetic testing (eg, in clinical or research settings), we think that the distinguishing factor is the primary recipient (ie, the consumers themselves and not clinicians) rather than the form of distribution or who actually pays for the tests. We also found several services (28 of 277) that offer genome sequencing to their consumers as opposed to only genotyping, potentially providing higher accuracy and clinical value than pure genotyping services. In consequence of these observations, our research supports the notion of an ongoing shift in the DTC genetic testing market to more mature DTC genomics 2.0.

Fourth, although most DTC genetic testing services were profit oriented, we found it interesting that there were at least 14 nonprofit services in our sample. Although these services did not form their own cluster in our 6-cluster solution, the majority of these services (n=9) can be found in cluster 5. They almost exclusively operate worldwide, only over the internet, and only offer third-party home collection kits and genotyping services. A notable exception to this is the DTC genetic testing service of Genes for Good that operates only in the United States, provides its own home collection kit and offers sequencing services. It is a nonprofit organization run by the University of Michigan, which seeks to engage people in genetic research. Generally, nonprofit DTC genetic testing services heavily rely on other for-profit DTC genetic testing services by, for example, using the test kits of other services or directly requesting consumers to import their data from other services. Specifically, the last scenario, where consumers freely and openly share their genome data with for-profit and nonprofit services, has received attention from researchers interested in socioeconomic perspectives of genome data sharing and crowdsourcing [1,24] and could be seen as an indicator for a trend toward *platformization* and a platform economy in genomics.

Limitations

Limitations of our study are as follows. First, the DTC genetic testing market is a volatile market with new services regularly appearing, mergers and acquisitions constantly happening, and extant services disappearing or changing their business strategies. Our taxonomy and archetypes, however, represent a snapshot of the current landscape of DTC genetic testing service business models. It is likely that in the meantime, new services would have emerged, while some services in our sample would have changed their business models or completely disappeared from the market. Nevertheless, we are confident that the developed taxonomy and derived archetypes build a strong foundation for further research in this area because of the rich and meaningful sample of DTC genetic testing services used and rigorous development of both. Moreover, owing to the extendibility ending condition being met, the presented taxonomy can easily be extended or altered in the case of the emergence of new services or additional insights. Second, our examination of services was mainly based on information retrieved from nonscientific sources and services' internet presences. Some information provided was ambiguous or not given at all. It is possible that our coding of some of the services are not entirely appropriate. We sought to address these information deficits by having services examined by 2 researchers independently, consulting additional internet sources, and, where no information was found, making informed guesses by comparing specific services with other similar services. Third, although Ward's method indicated the validity of several cluster solutions, analysis of common metrics provided inconclusive results on which cluster solution to choose for the *k*-means algorithm (ie, the dendrogram suggested 4, 5, and 6 cluster solutions, whereas the scree plot suggested 4, 5, 7, and 9 cluster solutions). After thorough examination of these cluster solutions based on a *k*-means clustering, we deemed the 6-cluster solution to be the most promising one for this research. Nevertheless, the 7-cluster solution might have provided additional insights into the DTC genomics market, which are not captured by the current cluster solution.

Implications and Future Research

Our study yields several implications for research and practice. For research, we provide a systematic classification of DTC genetic testing service business models and expose their essential characteristics. The developed taxonomy adds to our knowledge of business models in genomics, and in conjunction with the proposed archetypes, it contributes to a more comprehensive understanding of the DTC genetic testing industry. Compared with other taxonomies about technologies that are similarly disruptive as DTC genetic testing [35], the taxonomy presented here focuses on a rather narrow aspect of the DTC genetic testing service phenomenon. Although the

objective of this research was specifically to analyze an area of DTC genetic testing that has received little attention from researchers thus far (ie, the business models in this very industry), we also think that the presented taxonomy can serve as a starting point to analyze the DTC genetic testing phenomenon as a whole by, for example, broadening the taxonomy's scope and further developing it into a general taxonomy of DTC genetic testing services. However, we also note that despite the seminal work of Hwang and Christensen [8], the literature on the classification of business models in health care remains scarce. The presented taxonomy and archetypes can serve as an outset for new avenues in future research on DTC genetic testing. Starting from the previously outlined limitations of this study, researchers could seek to replicate the results of our research, especially considering the likely emergence of new DTC genetic services. Furthermore, as the taxonomy and archetypes only capture a snapshot of current DTC genetic testing service business models, future research should also attempt to analyze how those services' business models change over time. To this end, the literature on business model innovation [52,53] could provide a promising foundation to analyze the evolution of business models in DTC genetic testing. Although the clusters presented in this work were developed independent of any temporal dimension, future research may also investigate whether these clusters relate to different evolutionary stages of DTC genetic testing business models. From a socioeconomic and genetic privacy research perspective, our research provides a starting point to better understand the economic value of genomic data. We thereby support and strengthen the notion of information privacy as a commodity that can be traded in the context of genomics [54]. Specifically, dimensions in the categories such as strategic choices, create value, and capture value might prove useful to better understand the relationships between business models, genetic privacy, and crowdsourcing in genomics.

In terms of practice, our research has important implications for policy makers, professionals in the health care industry, DTC genetic testing services themselves, and consumers of such services. For policy makers, our research highlights a reality of diverse business models within the DTC genetic testing market. Considering DTC genetic testing services' need to collaborate closely with regulatory bodies, the developed taxonomy and archetypes can assist policy makers in designing policies that adapt to the diverse business landscape as to better protect consumers' well-being and privacy while respecting their right to informational self-determination. Similarly, the archetypes of DTC genetic testing service business models help raise awareness for the existence of different kinds of services with diverse benefits and risks for consumers and the health care system. Overall, our work contributes to a more nuanced understanding of DTC genetic testing in the health care sector, which has become progressively important given that evidence from studies among U.S. populations (one of the largest DTC genetic testing markets) suggests patients increasingly talk to their primary care physicians about DTC genetic testing [55,56], and expect them to be able to answer questions about DTC genetic test results [57]. For DTC genetic testing services, especially young services, our taxonomy serves as a valuable tool to analyze and possibly further develop their own business models as well as to analyze competitors' business models. The presented archetypes provide decision makers of DTC genetic testing services with blueprints of potential business models. Such blueprints can be used as initial guidance for transitioning from 1 business model to another or to identify market niches. From a consumer's point of view, it was a pleasant surprise that most DTC genetic testing service providers do not resell access to their consumers' genomic data to third parties. However, at the same time, our analysis reveals that many of the examined DTC genetic testing services retain the right to use collected genomic data for service improvements, something which consumers should be aware of. Overall and in line with research seeking to understand and improve consumers' understanding of DTC genetic testing [13,14], the taxonomy and archetypes presented in this study can serve consumers as tools for assessing DTC genetic testing services and finding services that best fit their needs.

Conclusions

DTC genetic testing is a relatively young and dynamic business area, which pushes genomics research forward and promises faster, as well as more affordable genomics services. Despite the rapid growth of the business sector, many concerns remain unanswered, and there is little knowledge about the impact of business models on DTC genetic testing services within research literature. In this study, we provide the first overview of DTC genetic testing service business models, resulting in a rigorously developed taxonomy and 6 service archetypes. This provides novel insights into the value and use of genomic data and can serve as a foundation for advanced research on relationships among DTC genetic testing services, their consumers, practitioners within the health care sector, and policy makers.

Acknowledgments

The authors acknowledge the support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of included and excluded direct-to-consumer genetic testing services.
[XLSX File (Microsoft Excel File), 61 KB - jmir_v22i1e14890_app1.xlsx]

Multimedia Appendix 2
Taxonomy development approach.
[PDF File (Adobe PDF File), 137 KB - jmir_v22i1e14890_app2.pdf]

Multimedia Appendix 3
Taxonomy development iterations details.
[PDF File (Adobe PDF File), 72 KB - jmir_v22i1e14890_app3.pdf]

Multimedia Appendix 4
Cluster analysis details.
[PDF File (Adobe PDF File), 107 KB - jmir_v22i1e14890_app4.pdf]

Multimedia Appendix 5
Taxonomy coding overview.
[XLSX File (Microsoft Excel File), 99 KB - jmir_v22i1e14890_app5.xlsx]

Multimedia Appendix 6
Cluster analysis details.
[PDF File (Adobe PDF File), 294 KB - jmir_v22i1e14890_app6.pdf]

References

1. Turrini M. Online Genomes: problematizing the disruptiveness of direct-to-consumer genetic tests. Sociol Compass 2018;12(11):e12633. [doi: 10.1111/soc4.12633]
2. Wetterstrand K. National Human Genome Research Institute. 2016. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) URL: https://www.genome.gov/sequencingcostsdata/ [accessed 2018-05-02] [WebCite Cache ID 6z7CocSFu]
3. Ramos E, Weissman SM. The dawn of consumer-directed testing. Am J Med Genet C Semin Med Genet 2018 Mar;178(1):89-97. [doi: 10.1002/ajmg.c.31603] [Medline: 29512889]
4. Credence Research. 2018. Global Direct-to-Consumer Genetic Testing Market is growing with Double Digit CAGR URL: https://www.credenceresearch.com/press/global-direct-to-consumer-genetic-testing-market [accessed 2018-05-02] [WebCite Cache ID 6z7BulybP]
5. Hudson K, Javitt G, Burke W, Byers P. ASHG statement on direct-to-consumer genetic testing in the United States. Am J Hum Genet 2007;81(3):635-637. [doi: 10.1086/521634]
6. Allyse MA, Robinson DH, Ferber MJ, Sharp RR. Direct-to-consumer testing 2.0: emerging models of direct-to-consumer genetic testing. Mayo Clin Proc 2018 Jan;93(1):113-120. [doi: 10.1016/j.mayocp.2017.11.001] [Medline: 29304915]
7. Dwivedi N. Bolly Zoom. 2019 Jan 6. How 23andMe Makes Money? Understanding 23andMe Business Model URL: https://www.bollyzoom.com/how-23andme-makes-money-understanding-23andme-business-model/ [accessed 2018-05-02] [WebCite Cache ID 6z7CgAPOC]
8. Hwang J, Christensen CM. Disruptive innovation in health care delivery: a framework for business-model innovation. Health Aff (Millwood) 2008;27(5):1329-1335. [doi: 10.1377/hlthaff.27.5.1329] [Medline: 18780919]
9. Thiebes S, Lyytinen K, Sunyaev A. Sharing is About Caring? Motivating and Discouraging Factors in Sharing Individual Genomic Data. In: Proceedings of the 38th International Conference on Information Systems. 2017 Presented at: ICIS'17; December 10-13, 2017; Seoul, South Korea p. 1-20.
10. Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. Annu Rev Genomics Hum Genet 2008;9:161-182. [doi: 10.1146/annurev.genom.9.081307.164319] [Medline: 18767961]
11. Bloss CS, Wineinger NE, Darst BF, Schork NJ, Topol EJ. Impact of direct-to-consumer genomic testing at long term follow-up. J Med Genet 2013 Jun;50(6):393-400. [doi: 10.1136/jmedgenet-2012-101207] [Medline: 23559530]
12. Leighton JW, Valverde K, Bernhardt BA. The general public's understanding and perception of direct-to-consumer genetic test results. Public Health Genomics 2012;15(1):11-21 [FREE Full text] [doi: 10.1159/000327159] [Medline: 21720148]
13. Shaer O, Nov O, Okerlund J, Balestra M, Stowell E, Ascher L, et al. Informing the design of direct-to-consumer interactive personal genomics reports. J Med Internet Res 2015 Jun 12;17(6):e146 [FREE Full text] [doi: 10.2196/jmir.4415] [Medline: 26070951]
14. Westendorf L, Shaer O, Pollalis C, Verish C, Nov O, Ball MP. Exploring genetic data across individuals: design and evaluation of a novel comparative report tool. J Med Internet Res 2018 Sep 24;20(9):e10297 [FREE Full text] [doi: 10.2196/10297] [Medline: 30249582]
15. Darnovsky M. The New York Times. 2015 Mar 2. 23andMe's Dangerous Business Model URL: https://www.nytimes.com/roomfordebate/2015/03/02/23andme-and-the-promise-of-anonymous-genetic-testing-10/23andmes-dangerous-business-model-17 [accessed 2018-05-02] [WebCite Cache ID 6z7ByDxxs]

16. Groberg J, Iqbal H, Tu E. UBS Evidence Lab. 2016 May 03. Genomics 2.0 – Consumer Genomics URL: https://www.lifenome.com/wp-content/uploads/2017/12/UBS-report.pdf [accessed 2018-05-02] [WebCite Cache ID 6z7CECBR1]
17. Shafer SM, Smith HJ, Linder JC. The power of business models. Bus Horiz 2005;48(3):199-207. [doi: 10.1016/j.bushor.2004.10.014]
18. Zott C, Amit R, Massa L. The business model: recent developments and future research. J Manag 2011;37(4):1019-1042. [doi: 10.1177/0149206311406265]
19. Osterwalder A, Pigneur Y, Tucci CL. Clarifying business models: origins, present, and future of the concept. Commun Assoc Inf Syst 2005;16:1-25. [doi: 10.17705/1cais.01601]
20. Herzlinger RE. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66, 156. [Medline: 16649698]
21. Covolo L, Rubinelli S, Ceretti E, Gelatti U. Internet-based direct-to-consumer genetic testing: a systematic review. J Med Internet Res 2015 Dec 14;17(12):e279 [FREE Full text] [doi: 10.2196/jmir.4378] [Medline: 26677835]
22. Caulfield T, Ries NM, Ray PN, Shuman C, Wilson B. Direct-to-consumer genetic testing: good, bad or benign? Clin Genet 2010 Feb;77(2):101-105. [doi: 10.1111/j.1399-0004.2009.01291.x] [Medline: 19968666]
23. Ducournau P, Gourraud PA, Rial-Sebbag E, Cambon-Thomsen A, Bulle A. Direct-to-consumerhealth genetic testing services: what commercial strategies for which socio-ethical issues? Health Sociol Rev 2013;22(1):2108-2137. [doi: 10.5172/hesr.2013.2108]
24. Riso B, Tupasela A, Vears DF, Felzmann H, Cockbain J, Loi M, et al. Ethical sharing of health data in online platforms - which values should be considered? Life Sci Soc Policy 2017 Aug 21;13(1):12 [FREE Full text] [doi: 10.1186/s40504-017-0060-z] [Medline: 28825221]
25. Levina M. Googling your genes: personal genomics and the discourse of citizen bioscience in the network age. J Sci Commun 2010;9(1):1-8. [doi: 10.22323/2.09010206]
26. Prainsack B. Understanding participation: the 'citizen science' of genetics. In: Schicktanz S, Werner-Felmayer G, Prainsack B, editors. Genetics as Social Practice: Transdisciplinary Views on Science and Culture. Farnham, UK: Ashgate Publishing, Ltd; 2014:147-164.
27. Borry P, Cornel MC, Howard HC. Where are you going, where have you been: a recent history of the direct-to-consumer genetic testing market. J Community Genet 2010 Sep;1(3):101-106 [FREE Full text] [doi: 10.1007/s12687-010-0023-z] [Medline: 21475669]
28. Wright CF, Gregory-Jones S. Size of the direct-to-consumer genomic testing market. Genet Med 2010 Sep;12(9):594. [doi: 10.1097/GIM.0b013e3181ead743] [Medline: 20697288]
29. Vanhala A, Reijonsaari K. Direct-to-Consumer Genome Data Services and Their Business Models. Helsinki: Sitra; 2013.
30. Remane G, Nickerson R, Hanelt A, Tesch JF, Kolbe LM. A Taxonomy of Carsharing Business Models. In: Proceedings of the 37th International Conference on Information Systems. 2016 Presented at: ICIS'16; December 11-14, 2016; Dublin, Ireland p. 1-19.
31. Phillips AM. Webpage archive. 2018 Feb 19. Data on Direct-to-Consumer Genetic Testing and DNA testing companies URL: http://archive.is/IDU5e [accessed 2019-09-09]
32. DNA Testing Choice. 2017. A-Z Tests URL: https://dnatestingchoice.com/tests [accessed 2018-05-02] [WebCite Cache ID 6z7C7DB8u]
33. International Society of Genetic Genealogy. 2017. List of DNA Testing Companies URL: https://isogg.org/wiki/List_of_DNA_testing_companies [accessed 2018-05-02] [WebCite Cache ID 6z7CHvaYk]
34. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. Eur J Inf Syst 2013;22(3):336-359. [doi: 10.1057/ejis.2012.26]
35. Gao F, Thiebes S, Sunyaev A. Rethinking the meaning of cloud computing for health care: a taxonomic perspective and future research directions. J Med Internet Res 2018 Jul 11;20(7):e10041 [FREE Full text] [doi: 10.2196/10041] [Medline: 29997108]
36. Thiebes S, Kleiber G, Sunyaev A. Cancer Genomics Research in the Cloud: A Taxonomy of Genome Data Sets. In: Proceedings of the 4th International Workshop on Genome Privacy and Security. 2017 Presented at: GenoPri'17; October 15, 2017; Orlando, Florida.
37. Schmidt-Kraepelin M, Thiebes S, Tran M, Sunyaev A. What's in the Game? Developing a Taxonomy of Gamification Concepts for Health Apps. In: Proceedings of the 51th Hawaii International Conference on System Sciences. 2018 Presented at: 51th Hawaii International Conference on System Sciences; January 3-6, 2018; Waikoloa Village, Hawaii p. 1217-1226. [doi: 10.24251/hicss.2018.150]
38. Mrosek R, Dehling T, Sunyaev A. Taxonomy of health IT and medication adherence. Health Policy Technol 2015;4(3):215-224. [doi: 10.1016/j.hlpt.2015.04.003]
39. Hors-Fraile S, Rivera-Romero O, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit-Balcells A, et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: a scoping review. Int J Med Inform 2018 Jun;114:143-155. [doi: 10.1016/j.ijmedinf.2017.12.018] [Medline: 29331276]
40. Schneider JA, Holland CP. eHealth search patterns: a comparison of private and public health care markets using online panel data. J Med Internet Res 2017 Apr 13;19(4):e117 [FREE Full text] [doi: 10.2196/jmir.6739] [Medline: 28408362]

41. Lavazova O, Dehling T, Sunyaev A. From Hype to Reality: A Taxonomy of Blockchain Applications. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: HICSS'19; January 8-11, 2019; Maui, Hawaii p. 4555-4564. [doi: 10.24251/hicss.2019.552]
42. Becker M, Matt C, Widjaja T, Hess T. Understanding Privacy Risk Perceptions of Consumer Health Wearables – an Empirical Taxonomy. In: Proceedings of the 38th International Conference on Information Systems. 2017 Presented at: ICIS'17; December 10-13, 2017; Seoul, South Korea p. 1-21.
43. Schneider S, Lansing J, Gao F, Sunyaev A. A Taxonomic Perspective on Certification Schemes: Development of a Taxonomy for Cloud Service Certification Criteria. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS'14; January 6-9, 2014; Waikoloa, Hawaii p. 4998-5007. [doi: 10.1109/hicss.2014.614]
44. Ge J, Gretzel U. A taxonomy of value co-creation on Weibo – a communication perspective. Int J Contemp Hosp Manag 2018;30(4):2075-2092. [doi: 10.1108/ijchm-09-2016-0557]
45. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, New Jersey: John Wiley & Sons; 2009.
46. Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. J Mark Res 1983;20(2):134-148. [doi: 10.2307/3151680]
47. Landau S, Everitt B. Analysis of repeated measures II: Linear mixed model. In: A Handbook of Statistical Analysis Using SPSS. New York: Chapman and Hall/CRC; 2004:194-215.
48. Hartmann PM, Zaki M, Feldmann N, Neely A. Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. Int J Oper Prod Manag 2016;36(10):1382-1406. [doi: 10.1108/ijopm-02-2014-0098]
49. Eickhoff M, Muntermann J, Weinrich T. What do FinTechs actually do? A taxonomy of FinTech business models. In: Proceedings of the 38th International Conference on Information Systems. 2017 Presented at: ICIS'17; December 10-13, 2017; Seoul, South Korea p. 1-19.
50. Imai K, Kricka LJ, Fortina P. Concordance study of 3 direct-to-consumer genetic-testing services. Clin Chem 2011 Mar;57(3):518-521 [FREE Full text] [doi: 10.1373/clinchem.2010.158220] [Medline: 21159896]
51. Stoeklé HC, Mamzer-Bruneel M, Vogt G, Hervé C. 23andMe: a new two-sided data-banking market model. BMC Med Ethics 2016 Mar 31;17:19 [FREE Full text] [doi: 10.1186/s12910-016-0101-9] [Medline: 27059184]
52. Chesbrough H. Business model innovation: opportunities and barriers. Long Range Plann 2010;43(2-3):354-363. [doi: 10.1016/j.lrp.2009.07.010]
53. Amit R, Zott C. Creating value through business model innovation. MIT Sloan Manag Rev 2012;53(3):41-49. [doi: 10.1002/9781118785317.weom030009]
54. Smith HJ, Dinev T, Xu H. Information privacy research: an interdisciplinary review. MIS Quarterly 2011;35(4):989-1016. [doi: 10.2307/41409970]
55. Kolor K, Liu T, St Pierre J, Khoury MJ. Health care provider and consumer awareness, perceptions, and use of direct-to-consumer personal genomic tests, United States, 2008. Genet Med 2009 Aug;11(8):595. [doi: 10.1097/GIM.0b013e3181b1cc2c] [Medline: 19680046]
56. Powell KP, Cogswell WA, Christianson CA, Dave G, Verma A, Eubanks S, et al. Primary care physicians' awareness, experience and opinions of direct-to-consumer genetic testing. J Genet Couns 2012 Feb;21(1):113-126. [doi: 10.1007/s10897-011-9390-9] [Medline: 21769569]
57. McGuire AL, Diaz CM, Wang T, Hilsenbeck SG. Social networkers' attitudes toward direct-to-consumer personal genome testing. Am J Bioeth 2009;9(6-7):3-10 [FREE Full text] [doi: 10.1080/15265160902928209] [Medline: 19998099]

Abbreviations

DTC: direct-to-consumer
FDA: US Food and Drug Administration

Edited by G Eysenbach; submitted 30.05.19; peer-reviewed by M Puttaroo, O Shaer; comments to author 26.08.19; revised version received 09.09.19; accepted 24.09.19; published 21.01.20.

Please cite as:
Thiebes S, Toussaint PA, Ju J, Ahn JH, Lyytinen K, Sunyaev A
Valuable Genomes: Taxonomy and Archetypes of Business Models in Direct-to-Consumer Genetic Testing
J Med Internet Res 2020;22(1):e14890
URL: https://www.jmir.org/2020/1/e14890
doi:10.2196/14890
PMID:31961329

Review

Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed

Jing Wang[1], MS; Huan Deng[1], MS; Bangtao Liu[1], MS; Anbin Hu[1], PhD; Jun Liang[2], MS; Lingye Fan[3], MS; Xu Zheng[4], MSc; Tong Wang[5], BS; Jianbo Lei[1,4,6], MD, PhD

[1]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

[2]IT Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[3]Affiliated Hospital, Southwest Medical University, Luzhou, China

[4]Center for Medical Informatics, Peking University, Beijing, China

[5]School of Public Health, Jilin University, Jilin, China

[6]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China

Corresponding Author:
Jianbo Lei, MD, PhD
Institute of Medical Technology
Health Science Center
Peking University
38 Xueyuan Rd
Haidian District
Beijing
China
Phone: 86 8280 5901
Email: jblei@hsc.pku.edu.cn

Abstract

Background: Natural language processing (NLP) is an important traditional field in computer science, but its application in medical research has faced many challenges. With the extensive digitalization of medical information globally and increasing importance of understanding and mining big data in the medical field, NLP is becoming more crucial.

Objective: The goal of the research was to perform a systematic review on the use of NLP in medical research with the aim of understanding the global progress on NLP research outcomes, content, methods, and study groups involved.

Methods: A systematic review was conducted using the PubMed database as a search platform. All published studies on the application of NLP in medicine (except biomedicine) during the 20 years between 1999 and 2018 were retrieved. The data obtained from these published studies were cleaned and structured. Excel (Microsoft Corp) and VOSviewer (Nees Jan van Eck and Ludo Waltman) were used to perform bibliometric analysis of publication trends, author orders, countries, institutions, collaboration relationships, research hot spots, diseases studied, and research methods.

Results: A total of 3498 articles were obtained during initial screening, and 2336 articles were found to meet the study criteria after manual screening. The number of publications increased every year, with a significant growth after 2012 (number of publications ranged from 148 to a maximum of 302 annually). The United States has occupied the leading position since the inception of the field, with the largest number of articles published. The United States contributed to 63.01% (1472/2336) of all publications, followed by France (5.44%, 127/2336) and the United Kingdom (3.51%, 82/2336). The author with the largest number of articles published was Hongfang Liu (70), while Stéphane Meystre (17) and Hua Xu (33) published the largest number of articles as the first and corresponding authors. Among the first author's affiliation institution, Columbia University published the largest number of articles, accounting for 4.54% (106/2336) of the total. Specifically, approximately one-fifth (17.68%, 413/2336) of the articles involved research on specific diseases, and the subject areas primarily focused on mental illness (16.46%, 68/413), breast cancer (5.81%, 24/413), and pneumonia (4.12%, 17/413).

Conclusions: NLP is in a period of robust development in the medical field, with an average of approximately 100 publications annually. Electronic medical records were the most used research materials, but social media such as Twitter have become

important research materials since 2015. Cancer (24.94%, 103/413) was the most common subject area in NLP-assisted medical research on diseases, with breast cancers (23.30%, 24/103) and lung cancers (14.56%, 15/103) accounting for the highest proportions of studies. Columbia University and the talents trained therein were the most active and prolific research forces on NLP in the medical field.

(J Med Internet Res 2020;22(1):e16816) doi:10.2196/16816

KEYWORDS

natural language processing; clinical; medicine; information extraction; electronic medical record

Introduction

Natural language processing (NLP) refers to the ability of machines to understand and explain the way humans write and talk. It involves studying various theories and methods that can realize effective communication between humans and computers in natural language and is an important direction in the field of artificial intelligence [1]. The goal of NLP is to realize human-like language understanding for a wide range of applications and tasks [2]. The earliest study on natural language understanding was the machine translation design first proposed by American Warren Weaver in 1949 [3].

In modern medical care, electronic health record (EHR) and electronic medical record (EMR) systems are undergoing rapid and large-scale development [4]. For example, in 2011, the Chinese government invested ¥630 million (US $97 million) to conduct a pilot project on primary medical and health care information systems for EHR, EMR, and outpatient management [5,6]. Medical records are valuable assets of hospitals that contain a large amount of important information, such as patients' chief complaints, diagnostic information, drugs administered, and adverse reactions. However, medical records have long been ineffectively used due to technological limitations and unstructured text formats [7]. NLP can transform these unstructured medical texts into structured data that contain important medical information from which scientists and medical personnel can identify useful medical data [8,9], thereby improving the quality and reducing the operating costs of the medical system. An increasing number of practical problems in medicine can now be solved using NLP, such as the detection of adverse drug reactions [10,11], information extraction from EHR [12], and EMR or EHR classification [13]. NLP can also be used to process issues in radiology research [14,15]. The use of NLP to aid the resolution of medical problems is advancing rapidly and drawing increasing attention [16].

With the rapid development of NLP in the medical field, there is a constant increase in the number of NLP-related articles, which has led to the accumulation of a substantial amount of research findings. Analyzing these articles can indirectly reflect the dynamic progress of NLP development in the medical field. Moreover, the results of the analysis can provide various benefits to academia, especially to scholars who are interested in pursuing careers in specific areas. Regarding the analysis and research, the studies by Cobo et al [17,18] define bibliometrics as the use of statistical methods for quantitative assessment of academic output. Bibliometrics is often used to discover top authors and institutions in a field [19], determine the structure of a research field [20], identify important topics [21], and mine research directions [22].

Previous studies have analyzed and summarized the applications of NLP in the medical field. For example, Chen et al [23] conducted a bibliometric analysis of the outcomes of NLP in medical research over 10 years from 2007 to 2016. The authors comprehensively discussed the current research status in the field, including the top authors and institutions. However, their study only analyzed 10 years of data and covered NLP research in all biomedical fields, not specifically medical research. In addition, details on the collaborative relationships between prolific authors and the diseases studied using NLP were not described. In 2015, Névéol et al [24] published a systematic review in which they focused on screening NLP methods that had been applied to clinical texts or clinical outcomes in the year of 2014 through searching bibliographic databases. In 2016, Névéol et al [25] summarized the outstanding papers on clinical NLP in the previous year. These studies mainly summarized recent research and presented a selection of the best papers published in the field of clinical NLP but lacked a comprehensive analysis of the use of NLP in the medical field.

Other previously published studies [23-26] have also summarized the role of NLP in medical research; however, they have essentially only summarized the basic characteristics, such as the number of published articles on NLP, author information, and keywords. Systematic analyses on other major features of NLP in the medical field, such as the collaboration among authors, popular research topics, and current status of the key diseases involved have not been conducted. Therefore, a systematic review spanning a longer period of time with more systematic and comprehensive analyses is necessary. This study differs from previous publications in the following aspects: first, bibliometrics was employed to review the relevant materials of medical NLP spanning nearly 20 years, which was the longest time span compared with previous studies; second, in addition to the analysis of certain basic characteristics as in previous studies, we used the VOSviewer tool version 1.6.10 (Centre for Science and Technology Studies, Leiden University) to perform cluster analyses on the relationships among authors and popular research topics. Third, we provided detailed discussion on multiple aspects of NLP, such as the diseases involved in NLP research and research tasks performed using NLP. In addition, to highlight the applications of NLP in the medical field that aligned more closely to clinical practice, we specifically excluded studies in the biomedical field, such as molecular biology, to provide more research reference materials for peers who conduct NLP research in the medical field.

Methods

Data Sources and Search Strategies

PubMed is an important search engine. The source of the PubMed database is MEDLINE, and the core topic is medicine. The objective of this study was to collect academic articles on the application of NLP in medicine. Therefore, PubMed was selected as the search engine in this study. On the PubMed platform, the search strategy was ("natural language processing" [all fields] OR NLP [all fields]) AND (medical [all fields] OR health [all fields] OR clinical [all fields]), automatically translated by PubMed to: (("natural language processing" [MeSH terms] OR ("natural" [all fields] AND "language" [all fields] AND "processing" [all fields]) OR "natural language processing" [all fields]) OR NLP [all fields]) AND (medical [all fields] OR ("health" [MeSH terms] OR "health" [all fields]) OR clinical [all fields]), and the time period spanned from 1999 to 2018.

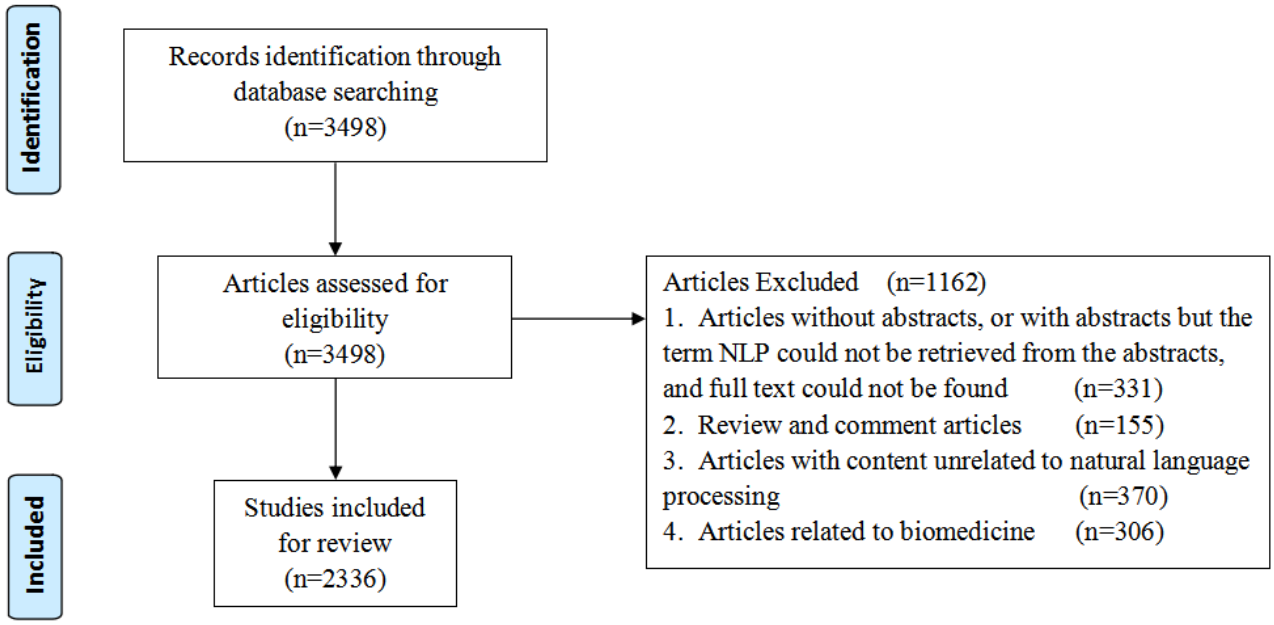
Inclusion and Exclusion Criteria

All published studies on the application of NLP in medicine (except biomedicine) during the 20 years between 1999 and 2018 were retrieved. A total of 3498 articles were retrieved. The articles were screened according to the following exclusion criteria:

- Articles with indeterminate content were excluded, including PubMed articles without abstracts and articles with abstracts but the term NLP could not be retrieved from the abstracts and the full text could not be found.
- Review and comment articles were excluded.
- Articles with content unrelated to NLP were excluded; for example, articles wherein the term NLP did not stand for natural language processing but for terms such as neurolinguistic programming, no light perception, and ninein-like protein or NLP was only mentioned as a previous study or future study, while the main article was unrelated to NLP.
- As the subject of this study was the application of NLP in medicine and diseases, articles on molecular biomedicine, such as studies on protein-protein interactions in biomedical studies [27], were excluded.

The first three steps of the screening process were mainly completed by JW, and the last step of screening was jointly completed by JW and HD. In cases of discordance during the screening process on whether the article belonged to the molecular biomedical category, the two authors would review the full text and come to an agreement through discussion. We followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [28], shown in Figure 1, for the screening procedure. A total of 2336 articles were included in the statistical analysis.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram depicting the screening procedure for articles on natural language processing (NLP) in the medical field.



Data Extraction and Statistical Analysis

The following information was extracted from eligible articles: year of publication, journal name in which the article was first published, all authors, first author, corresponding author, first author's affiliation institution (and department), first author's country, research tasks of NLP in the article, and disease type discussed in the article. The obtained data were input into Excel 2016 (Microsoft Corp) for data analysis and processing. Excel and VOSviewer were used in this study for the qualitative and quantitative analyses of author co-occurrences, keywords, and disease types, which helped compile and summarize the characteristics of the development of the medical NLP field in detail. The cutoff date for data collection was December 31, 2018.

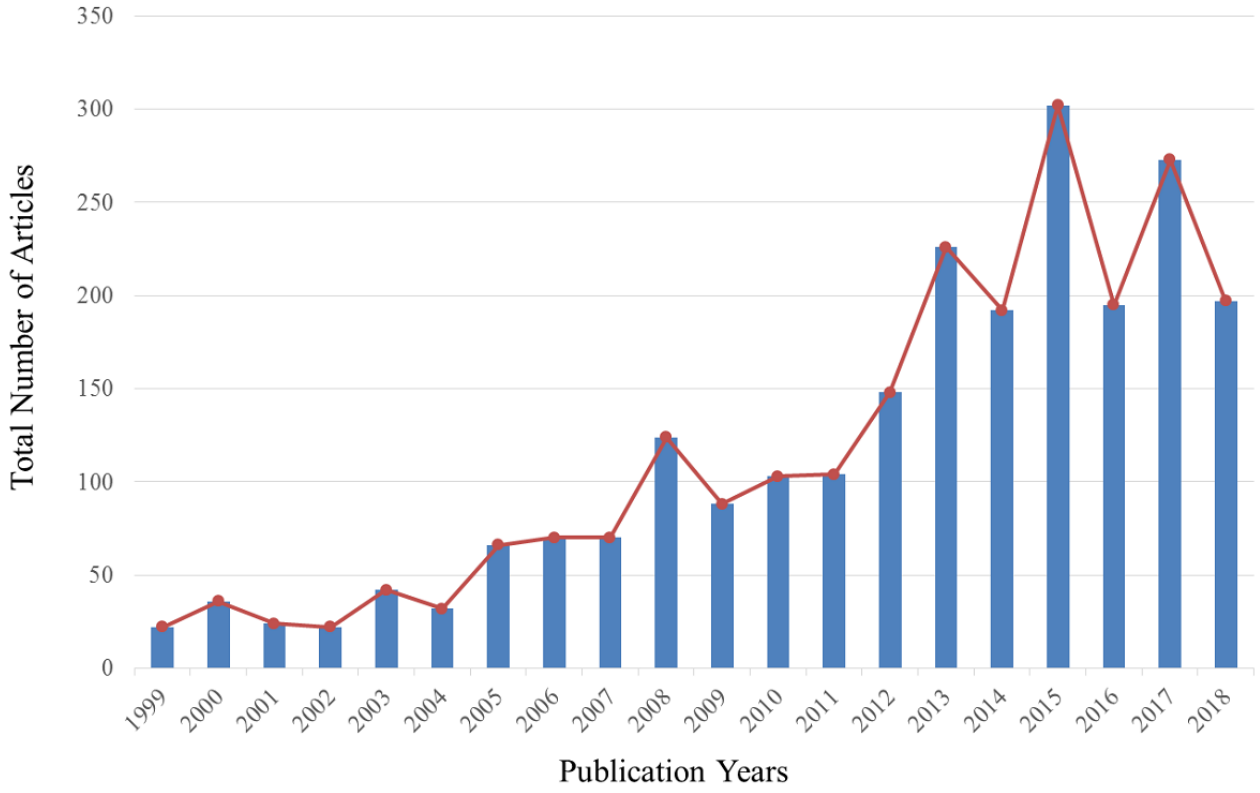
Results

Overall Analysis of Article Data

Trends in Number of Articles

Of the 2336 articles that met the study criteria, the time period spanned from 1999 to 2018. The overall trend (Figure 2) showed that the number of published articles increased every year. The time period was mainly divided into 3 phases: between 1999 and 2004 was the lag period, in which the development of the field was relatively slow, with an average of 30 (22 to 42) articles published; between 2005 and 2011 was the slow growth period, with an average of 89 (66 to 124) articles published; after 2012, NLP in the medical field entered a fast growth period. Until 2018, a yearly average of 219 (148 to 302) articles were published, with the peak (302) attained in 2015.

Figure 2. Graph showing the number of articles published over time.



Journals in Which Articles Were Published

A total of 2336 articles were published in 412 journals. Table 1 shows the names of the top 10 journals and the corresponding number of articles in each journal. These 10 journals together contained more than 50% of the total number of articles.

Table 1. Medical natural language processing journal rankings (n=2336).

Rank	Journal or proceedings	Publications, n (%)
1	Studies in Health Technology and Informatics	408 (17.47)
2	AMIA Annual Symposium Proceedings	386 (16.53)
3	Journal of the American Medical Informatics Association	256 (10.96)
4	Journal of Biomedical Informatics	223 (9.55)
5	International Journal of Medical Informatics	54 (2.31)
6	BMC Medical Informatics and Decision Making	50 (2.14)
7	BMC Bioinformatics	43 (1.84)
8	AMIA Joint Summits on Translational Science Proceedings	31 (1.33)
9	Plos ONE	31 (1.33)
10	Journal of Digital Imaging	30 (1.28)

Analysis of Author-Related Data

Author Orders

This study screened for the first author, corresponding author, and contributing authors of each article. The top 10 authors in each category are presented in Table 2 and Table 3. Specifically, Hongfang Liu, Hua Xu, and Joshua C Denny were ranked as the top three authors with the most number of articles published. The top three first authors were Stéphane Meystre, Özlem Uzuner, and Hua Xu, and the top three corresponding authors were Hua Xu, Stéphane Meystre and Özlem Uzuner and Carol Friedman (tie). There were four authors whose names appeared top 10 in each of the three categories: Hua Xu, Joshua C Denny, Wendy W Chapman, and Özlem Uzuner.

Table 2. Rank of top authors by number of articles published and the most articles published as the first plus corresponding author.

Total (first + corresponding + coauthor)			Total (first + corresponding)	
Rank	Authors	Publications	Publications	Rank
1	Hongfang Liu	70	21 (7+14)	6
2	Hua Xu	66	48 (15+33)	1
3	Joshua C Denny	64	26 (12+14)	4
4	Carol Friedman	60	20 (6+14)	7
5	Wendy W Chapman	55	25 (11+14)	5
6	Guergana Savova	45	—	—
6	Christopher G Chute	45	—	—
8	Serguei Pakhomov	43	—	—
9	Özlem Uzuner	37	—	—
9	George Hripcsak	37	—	—
9	Thomas C Rindflesch	37	—	—
—	Stéphane Meystre	—	32 (17+15)	2
—	Özlem Uzuner	—	30 (16+14)	3

Table 3. Top first authors and corresponding authors.

Author designation	Rank	Publications
First		
Stéphane Meystre	1	17
Özlem Uzuner	2	16
Hua Xu	3	15
Louise Deleger	4	13
Joshua C Denny	5	12
Serguei Pakhomov	5	12
Wendy W Chapman	7	11
Sunghwan Sohn	8	10
Li Zhou	9	9
Guergana Savova	9	9
Corresponding		
Hua Xu	1	33
Stéphane Meystre	2	15
Özlem Uzuner	3	14
Carol Friedman	3	14
Hongfang Liu	3	14
Wendy W Chapman	3	14
Joshua C Denny	3	14
Imre Solti	8	11
Genevieve B Melton	9	10
Hong Yu	9	10

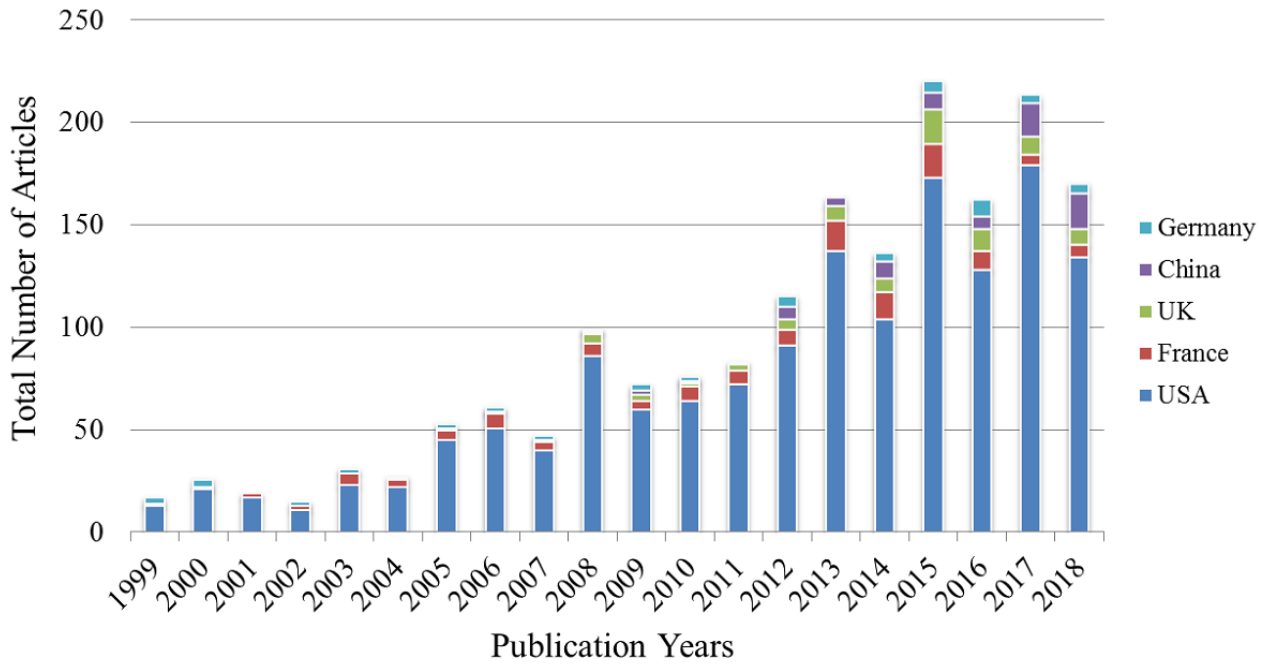
Countries in Which Authors Were Based

This study first analyzed the countries in which the first authors' institutions were located. The top 10 countries and the articles published are listed in Table 4, which shows that the United States is the top country and has contributed more than 50% of the total number of articles (63.01%), followed by France (5.44%), the United Kingdom (3.51%), and China (3.04%). Furthermore, in 2015 and 2017, the United States stood out with more than 150 articles published. Next, we analyzed the trend in the number of articles published in the top five countries over 20 years (Figure 3).

Table 4. Ranking of the first author's countries (top 10, n=2336).

Rank	Country	Publications, n (%)
1	United States	1472 (63.01)
2	France	127 (5.44)
3	United Kingdom	82 (3.51)
4	China	71 (3.04)
5	Germany	57 (2.44)
6	Australia	56 (2.40)
7	Japan	52 (2.23)
8	Switzerland	44 (1.88)
9	Canada	33 (1.41)
10	Spain	28 (1.20)

Figure 3. Trend in the number of articles published over 20 years in the top five countries with the most articles published.



Institutions to Which Authors Belonged

This study analyzed the relevant data on the institutions from which the articles were published. Specifically, the primary institutions to which the first authors belonged were analyzed (Table 5). The data showed that the top three institutions were Columbia University (4.54%), University of Utah (4.15%), and Mayo Clinic (3.85%). Together, these three institutions contributed a total of 12.54% of the articles published.

Table 5. Ranking of institutions to which the first authors belonged (n=2336).

Rank	Institution name	Publications, n (%)
1	Columbia University	106 (4.54)
2	University of Utah	97 (4.15)
3	Mayo Clinic	90 (3.85)
4	Vanderbilt University	59 (2.53)
5	National Library of Medicine	57 (2.31)
6	Brigham and Women's Hospital	52 (2.24)
7	University of California	47 (2.01)
8	University of Pittsburgh	38 (1.63)
9	Massachusetts General Hospital	37 (1.58)
10	University of Minnesota	32 (1.37)

Departments to Which Authors Belonged

This study evaluated the professional background of the first authors and analyzed the departments to which the first authors belonged, with the aim of observing the overall development of NLP in the medical field across the broad range of the discipline. As statistical analysis of institutions in this study focused on the primary institutions to which the authors belonged, analysis of departments also focused on departments of the primary institutions. If an author was affiliated to multiple departments, all departments were included in the statistical analysis. Table 6 shows that the top four departments are biomedical informatics (14.3%), computer science (6.0%), radiology (3.2%), and medical informatics (2.4%).

Table 6. Distribution of departments to which the first authors belonged (n=2336).

Rank	Name of department	Publications, n (%)
1	Department of biomedical informatics	334 (14.30)
2	Department of computer science	141 (6.04)
3	Department of radiology	75 (3.21)
4	Department of medical informatics	55 (2.35)
5	Department of psychiatry	37 (1.58)
6	Department of neuroscience	35 (1.50)
7	Department of nursing	30 (1.28)
8	Department of health sciences	28 (1.20)
9	Department of medicine	22 (0.94)
10	Department of health informatics	19 (0.81)

Collaboration Status Among Authors

VOSviewer is a bibliometric analysis software for constructing and visualizing bibliometric maps. It was codeveloped by Nees Jan van Eck and Ludo Waltman of Leiden University in the Netherlands [29], and it has unique advantages in clustering techniques based on co-occurrences. VOSviewer provides three types of map visualizations: network visualization, overlay visualization, and density visualization. VOSviewer was used in this study to analyze the collaboration status among authors, and the network visualization and overlay visualization of VOSviewer were employed. The network visualization could provide clusters of top authors in the field. This, together with the overlay visualization, could provide the distribution of timing of collaboration in each author cluster to understand their collaboration trends. The directions of collaboration and research objectives of each author cluster could then be obtained through reviewing the corresponding articles. When performing analysis using VOSviewer in this study, the minimum number of documents of an author was set to 20. As shown in Figure 4A, the article authors were divided into six large clusters, and Figure 4B shows the distribution of collaboration time among the authors.

Keyword Analysis

Analysis of keywords can indirectly reveal hotspots and changing trends in research topics, critical for understanding the development of this field [30]. VOSviewer was used in this study to perform keyword analysis. The purpose of the analysis was to identify the most popular research hotspots in the field and obtain the changing trends in keywords over time through the overlay visualization generated in VOSviewer. This could help researchers determine potential future research directions. During statistical analysis, keywords were defined as words that were used more than 50 times in titles and abstracts in all publications. As shown in Figure 5A, 327 keywords were identified, and the keywords were grouped as red, yellow, and blue. Based on these three categories, the relatedness among these keywords can be observed. For example, in the red category, patient (978 times), electronic health record (610 times), and electronic medical record (361 times) belong to the clinical NLP field; in the blue category, classifier (249 times), machine learning (215 times), support vector machine (164 times), and information extraction (150 times) belong to NLP research methods; and in the green category, language (449 times), phrase and word (395 times), ontology (345 times), terminology (267 times), and lexicon (106 times) belong to NLP research subjects. Next, the overlay visualization (Figure 5B) shows the trends in keyword changes as time progresses. In Figure 5B, blue indicates that the timing of appearance is earlier, and red indicates that the timing of appearance is later. The figure reveals certain hotspots have developed in the field in recent years, including electronic health record (176 times in 2014), cancer (19 times in 2014), and machine learning (34 times in 2014). It is worth noting that social media in the red category appeared 22 times in 2016.

Figure 4. (A) Network visualization of author co-occurrences analyzed using VOSviewer. A circle represents an author, the size of the circle represents the importance, and the thickness of the link connecting the circles represents the relatedness of the connections. Circles with the same color belong to the same cluster. (B) Overlay visualization generated in VOSviewer (Centre for Science and Technology Studies, Leiden University). A color closer to blue represents an earlier time and closer to red represents a time closer to 2018 (note: refer to Multimedia Appendix 1 for details on the two diagrams and related discussions).

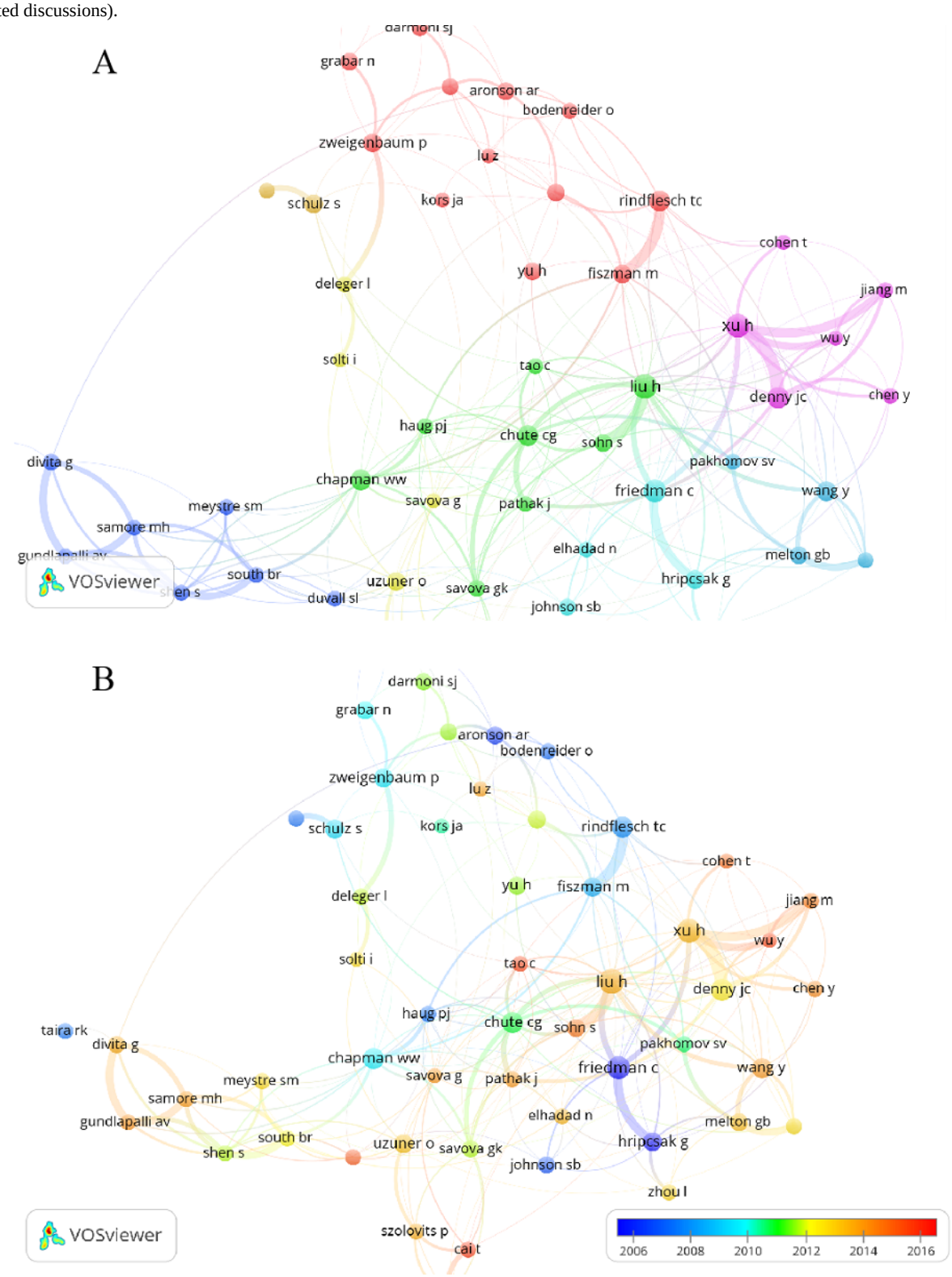
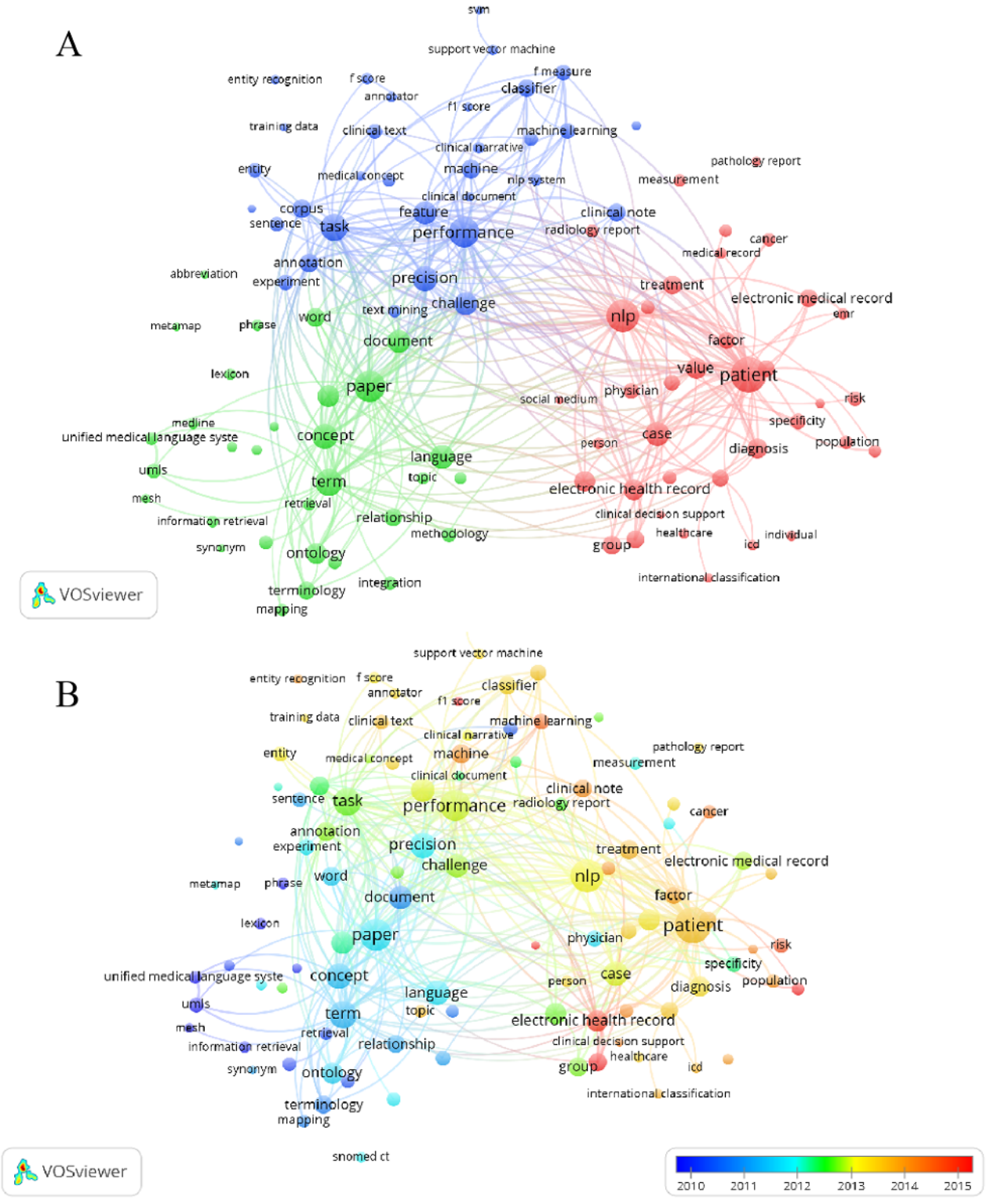


Figure 5. (A) Distribution of keywords. A circle represents an identified keyword, the size of the circle represents the importance, and the thickness of the link connecting the circles represents the relatedness of the connections among the keywords. Circles with the same color belong to the same cluster. (B) Changes in keywords over time. A color closer to blue represents an earlier time and closer to red represents a time closer to 2018 (note: refer to Multimedia Appendix 1 for details on the two diagrams and related discussions).



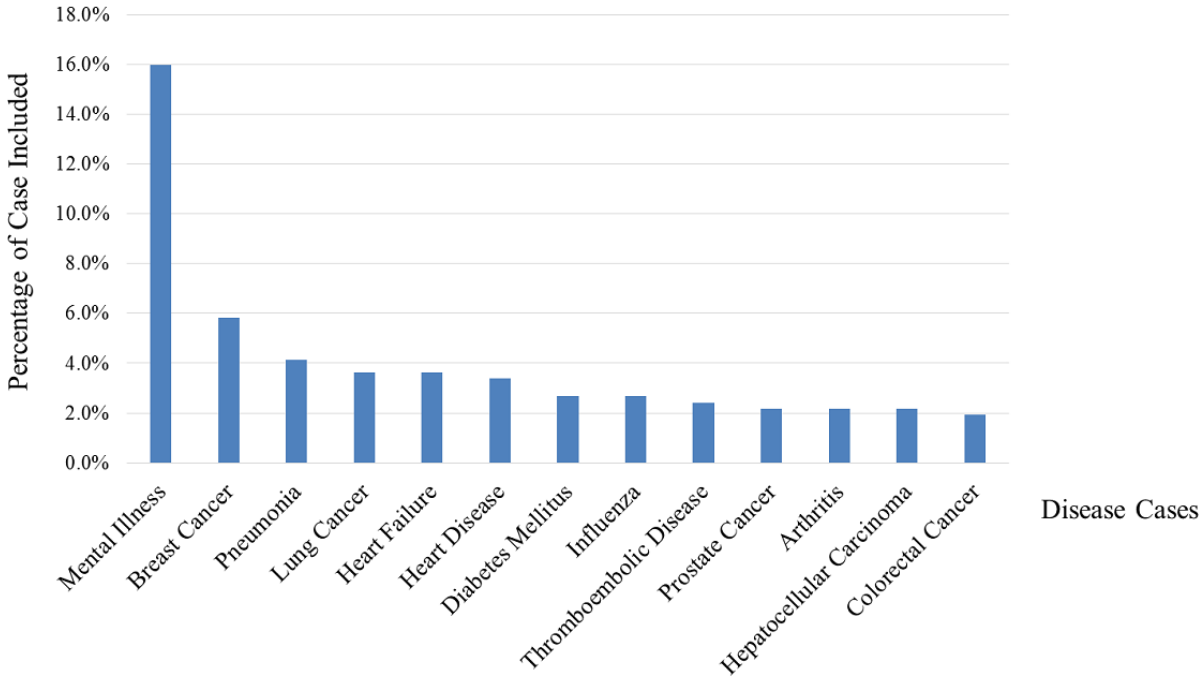
Analysis of Current Status of Specific Diseases Studied Using Natural Language Processing

This study found that 413 articles mentioned specific diseases studied using NLP, accounting for about one-fifth of the total number of articles. We conducted a comprehensive analysis of these articles to understand the type of disease information mined by NLP and how it was performed. This could provide a reference tool for the use of NLP when studying disease cases in the future.

Current Status of Specific Diseases Studied Using Natural Language Processing

Of the 413 articles, the categories of diseases studied using NLP are shown in Figure 6. Specifically, mental illness ranked at the top, accounting for 16.5% (68/413) of the articles. The second and third ranks were breast cancer (5.8%, 24/413) and pneumonia (4.1%, 17/413). The names of the diseases in the Figure 6 were mainly based on the specific disease names mentioned in the article.

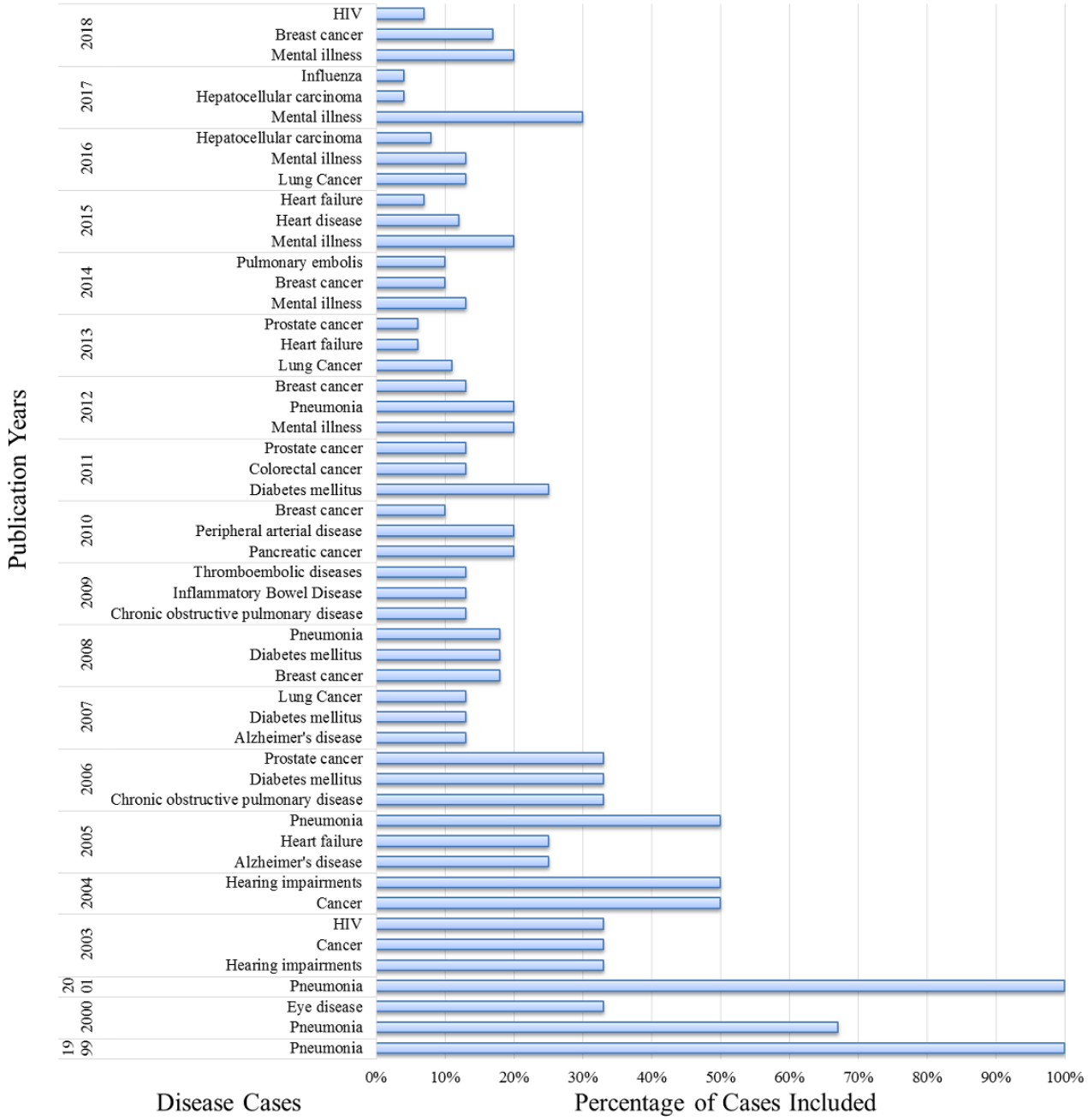
Figure 6. Ranking of disease categories based on studies that used natural language processing for the investigation of disease cases.



Specific Diseases Studied Using Natural Language Processing by Time Period

The temporal distribution of NLP research used to study diseases was analyzed in this study. As shown in Figure 7, initially in 1999, only one article clearly stated the type of disease that involved the use of NLP: pneumonia. In the next 3 years, pneumonia remained the main subject area in NLP research. From 2006, the use of NLP for the study of cancer cases had become popular, with a primary focus on lung cancer, prostate cancer, and breast cancer. The use of NLP in breast cancer research was mainly concentrated in 2018, with 10 articles published, almost all of them were from the United States. In addition, diseases such as diabetes, mental illness, and prostate cancer were all common subject areas in NLP research.

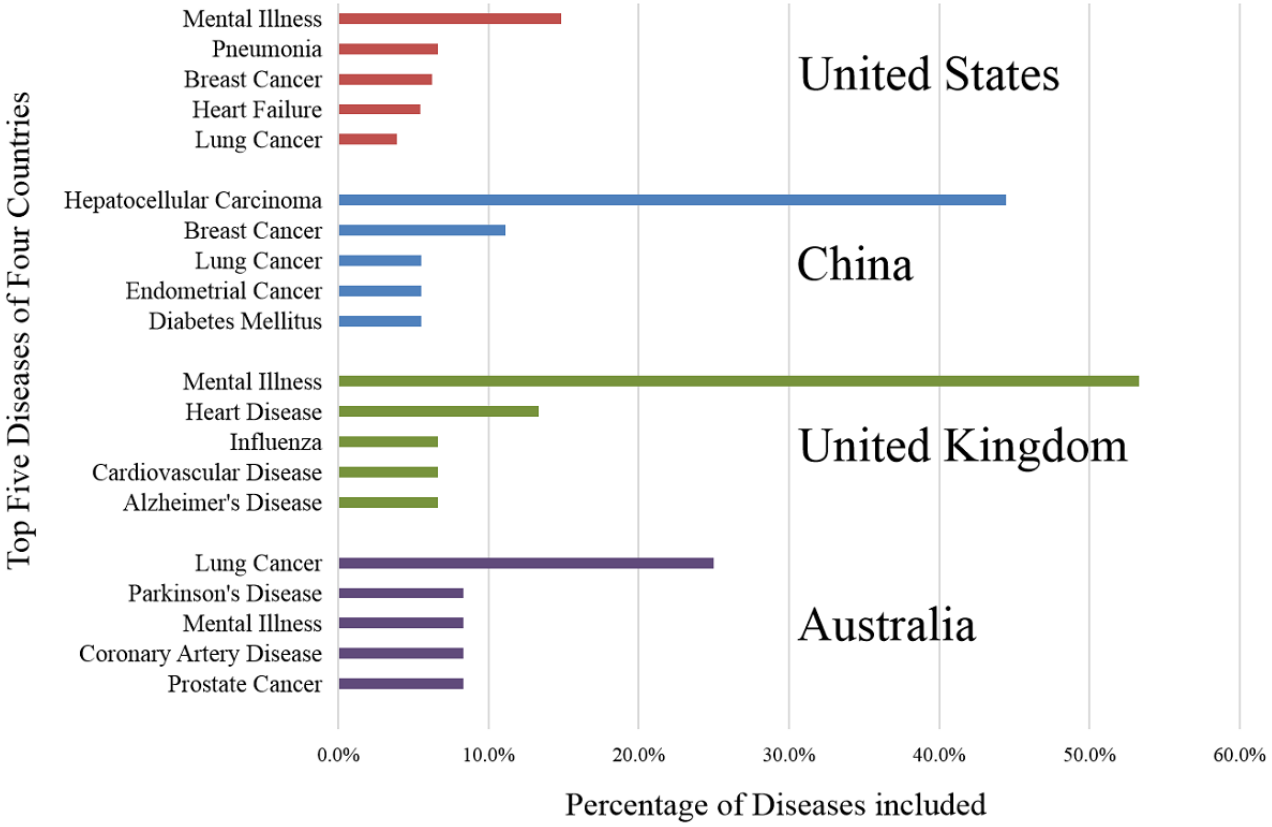
Figure 7. Temporal distribution of studies that used natural language processing for the investigation of disease cases (note: this figure shows the names of the top three diseases in studies that used natural language processing to investigate disease cases each year. Fewer than three disease types indicates that only one or two diseases were studied in the year. The term cancer in the figure indicates the article only mentioned the term cancer, without specifying the type of cancer).



Current Status of Diseases Studied Using Natural Language Processing by Country

Of the 413 articles that studied disease cases using NLP, the top four countries from where the first authors were located were the United States (68.3%, 282/413), China (4.8%, 20/413), the United Kingdom (3.6%, 15/413), and Australia (3.1%, 13/413). This ranking was consistent with the total number of articles published by country. The status of NLP research for use to study disease cases in these four countries was further investigated. As shown in Figure 8, the research subjects in the United States were more diverse, and there was no specific area of focus. The key subject area studied in China was hepatocellular carcinoma. The United Kingdom and Australia mainly focused on mental illness and lung cancer.

Figure 8. Distribution of diseases in studies that used natural language processing for the investigation of disease cases in the United States, China, United Kingdom, and Australia.

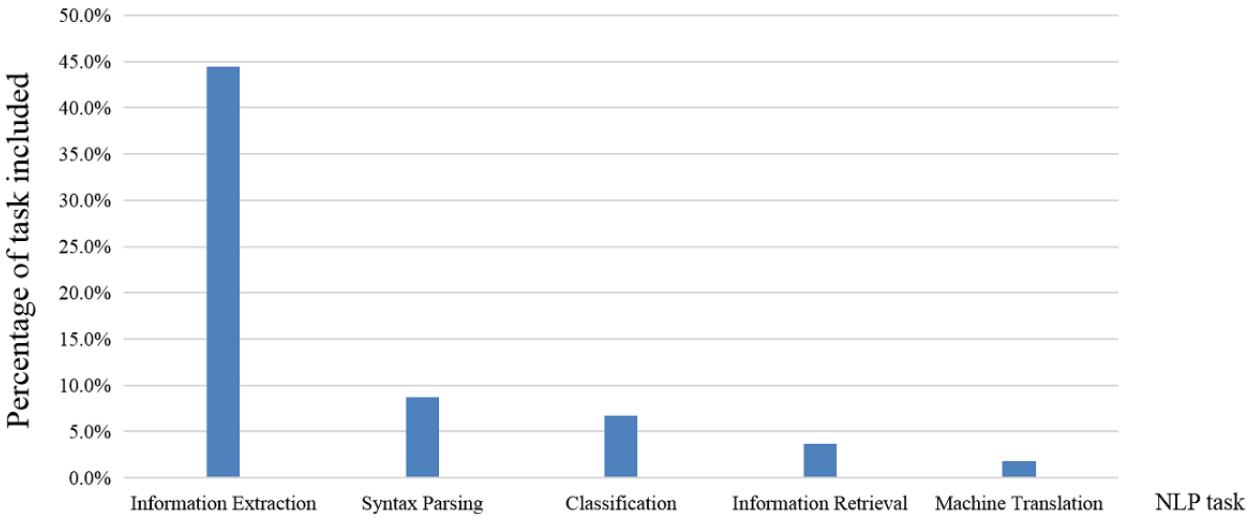


Research Tasks of Natural Language Processing in the Medical Field

The abstracts of 2336 articles were analyzed in this study to explore the research tasks of NLP involved in each article. If the abstract did not mention the specific task of NLP, the full text was reviewed. If the task could not be clearly identified from the full text, the article would be excluded from the analysis. NLP tasks involved were undetermined in 73 articles.

The authors of this study referenced the content on NLP described in chapter 4 of *Artificial Intelligence and its Application, Fourth Edition* [31], and divided the NLP tasks into speech recognition, machine translation, syntax parsing, classification, information retrieval, information extraction, information filtering, natural language generation, sentiment analysis, question answering system, and so on. This study analyzed the number of articles related to each NLP task and found that the top five tasks were information extraction (44.41%, 1005/2263), syntax parsing (8.66%, 196/2263), classification (6.72%, 152/2263), information retrieval (3.71%, 84/2263), and machine translation (1.77%, 40/2263; Figure 9).

Figure 9. Top five ranks of the research tasks of natural language processing (NLP) in the medical field.



Discussion

Overall Development Status of Medical Natural Language Processing

NLP research in the past 20 years could be divided into 3 phases: the lag period (1999-2004) with a yearly average of 30 (22 to 42) articles published, the slow growth period (2005-2011) with a yearly average of 89 (66 to 124) articles published, and the fast growth period (2012-2018) with a yearly average of 219 articles (148 to 302) articles published, with a peak (302) attained in 2015. Analysis by country showed that the United States has been the leader since the beginning of NLP development. Prior to 2008, only the United States, France, and Germany, with few exceptions, had conducted investigations in the field. Of the five countries shown in Figure 3, China started the latest and only began to emerge in the field in 2012. The development of NLP in Germany has remained relatively stable without a particular outstanding year, and Germany generally ranked in the fourth or fifth position. The development of NLP in France has also been relatively stable. In the first 15 years, France usually occupied the second position, but it has been surpassed by China in the past 2 years. Between 2016 and 2018, China has published nearly 40 articles, with a primary focus on hepatocellular carcinoma research assisted by NLP, as well as the use of NLP to mine or identify relevant information in clinical notes or EMR.

Analysis of Prolific Authors and Affiliation Institutions

This study identified the prominent authors who had made significant contributions to the NLP field, and we noted the following salient feature: the top two authors with the highest number of publications, Hongfang Liu and Hua Xu, plus Carol Friedman (ranked fourth rather than first because quite a few of her articles are about methodology and biology, which were not included in the scope of this study, but this does not change that she is recognized as a leading pioneer in this field) and George Hripcsak, ninth position, were all from Columbia University. In particular, Carol Friedman and George Hripcsak are currently at Columbia University, whereas Hongfang Liu and Hua Xu are both students of Carol Friedman. Among the top five prolific authors who published as the first plus corresponding author, Hua Xu (ranked first), Hongfang Liu (ranked sixth), and Carol Friedman (ranked seventh), were all from Columbia University. In addition, analysis of the first author's affiliation institutions showed that Columbia University (106) was ahead of University of Utah (97) in second place and the Mayo Clinic (90) in third place. These findings indicated that Columbia University and its students were the most active in the field of medical NLP research.

Notably, as shown in Table 3, the top 10 institutions to which the first authors belonged were all from the United States, including 6 universities, 3 hospitals, and 1 library. This also reflects that universities are the key locations for conducting medical NLP research.

Analysis by department showed that the top four majors were biomedical informatics, computer science, radiology, and medical informatics. These four majors mainly involve the processing of highly integrated data using computers and the expertise involved related to interdisciplinary content, such as medical information. It was evident that researchers with professional backgrounds in these fields had contributed significantly to the development of NLP. The research and study of NLP should be the key learning direction for future students majoring these subjects.

Current Development Status of Natural Language Processing Research on Disease Investigations

Analysis of this study showed that the top disease type in disease research involving NLP was mental illness. The World Health Organization predicts that mental illness may become the third most common human disease in the world in the future, after heart disease and cancer [32], showing the severity of the risk posed by this illness. NLP plays an indispensable role in mental illness research. For example, Victor et al [33] used NLP to train a diagnostic algorithm with 95% specificity for classifying bipolar disorder. It has been shown that NLP of EHRs is increasingly being used to study mental illness [34].

The journal Lancet Oncology published global cancer statistics for young people aged 20 to 39 years in 2017: one million young people in the world are diagnosed with cancer each year, and breast cancer is the most commonly diagnosed cancer (20%) [35]. Faced with such severe circumstances, Zeng et al [36] used NLP to investigate challenging issues in breast cancer such as local recurrence.

From 1999 to 2005, NLP was often used to study pneumonia cases. Our analysis showed that the main role of NLP in studies on pneumonia cases was the identification of pneumonia-related concepts from chest radiograph reports, or the use of NLP to complete automatic coding of pneumonia-related concepts. In addition, Jones et al [37] used a natural language processing tool to identify patients for pneumonia across US Department of Veterans Affairs emergency departments. The additional assistance provided by NLP improved physicians' ability to identify pneumonia and facilitated clinical decision making by physicians.

Among disease research involving NLP, China ranked second regarding the number of articles published (20 articles). Figure 8 shows that half the studies conducted by Chinese researchers exploring diseases using NLP are on hepatocellular carcinoma. Hepatocellular carcinoma is a primary liver cancer with a high mortality rate. Research on hepatocellular carcinoma in China was concentrated in 2016 and 2017. The research direction was mainly in two areas: (1) information extraction using NLP for mining relevant data [38] and (2) combining NLP analysis with other analyses, such as pathway analysis and ontology analysis, to mine the role of related genes in hepatocellular carcinoma, such as microRNA-132 and microRNA-223-3p [39].

Research Tasks of Natural Language Processing in Medicine

According to the results of this study, and as shown in Figure 9, the most widely performed tasks by NLP in the medical field were information extraction, syntax parsing, classification, information retrieval, and machine translation. We will now discuss these five tasks in detail.

Information extraction accounted for the highest proportion of all medical NLP tasks. Almost one-third of medical NLP tasks were information extraction, indicating its importance in NLP. Information extraction mainly refers to the use of computers to automatically extract a specific type of information (such as entities, relationships, and events) from a vast number of structured or semistructured texts and to form structured data [40]. The analysis in this study, together with a previously published report [40], concludes that the development of information extraction in the medical field includes four main parts: (1) entity recognition, in which the task is to identify content such as a person's name, time, and place from the texts and add the corresponding labeling information [41-44]; (2) anaphora resolution, which mainly refers to the way of simplifying and standardizing the expression of entities that can greatly improve the accuracy of the results from information extraction [45]; (3) relationship extraction, which obtains the grammatical or semantic connections among entities in the texts, such as temporal relationships and is a crucial element in information extraction [46,47]; and (4) event extraction, which mainly focuses on how to extract events of interest from unstructured texts containing event information and present the events expressed in natural language in a structured form [48-50]. The paper found that the platform of information extraction has gradually moved to social media; 20% of the articles obtained data through the Twitter platform [51-55].

Text classification, which is a process of automated text classification based on text content and the use of computers to automatically classify texts under a given classification system and classification criteria [31]. There were many cases involved text classification [56-58], for example, Morioka et al [56] developed a feature vector to classify the radiology reports with a decision table classifier.

Syntactic analysis, also known as parsing in natural language, uses syntax and other relevant knowledge of natural languages to determine the functions of each component that constitutes an input sentence. This technology is used to establish a data structure and acquire the meaning of the input sentence [31]. The process includes lexical analysis [59], grammatical analysis, and semantic analysis.

Information retrieval refers to the query methods and processes for searching related documents required by users from an enormous number of documents using computer systems [31]. For example, Tang et al [60] investigated a novel deep learning–based method to retrieve the similar patient question in Chinese.

Machine translation refers to the automated translation of words or speech from one natural language to another natural language using computer programs. To put in simple terms, machine translation is the conversion of words from one natural language into words of another language. More complex translations can be automated using corpora [31]. For example, Merabti et al [61] translated the Foundational Model of Anatomy terms into French using methods lexically based on several NLP tools.

Conclusions

In this study, we conducted a bibliometric analysis and presented the development of NLP in the medical field over the past 20 years. While the United States continues to be the leader in the field, many countries such as China and the United Kingdom are also advancing rapidly. In recent years, the use of NLP has become popular to process information obtained from social media platforms—for example, studies have obtained information related to diseases and patient care from the Twitter platform. Cancer has always been one of the greatest threats to human health. The use of NLP to assist cancer research has become a recent trend, for example, for use in breast cancer and prostate cancer research. Tasks such as information extraction and syntax parsing have always been popular tasks in the medical NLP field. Future studies will focus on how to better integrate these tasks into medical NLP research.

Acknowledgments

This study was sponsored by the National Natural Science Foundation of China (grants #81771937 and #81871455).

Authors' Contributions

JL developed the conceptual framework and research protocol for the study. JW and HD conducted the publications review, data collection, and analysis. BL, AH, TW, XZ, and JL interpreted the data, LF made sure the diseases were classified correctly. JW drafted the manuscript, and JL made major revisions. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Network diagrams and analysis of keywords and collaboration among authors.
[DOCX File , 1678 KB - jmir_v22i1e16816_app1.docx]

References

1. Cambria E, White B. Jumping NLP curves: a review of natural language processing research [review article]. IEEE Comput Intell Mag 2014 May;9(2):48-57. [doi: 10.1109/mci.2014.2307227]

2. Liddy E. Natural language processing. Scripting Intelligence 2001;10(1):450-461. [doi: 10.1007/978-1-4302-2352-8_3]
3. Weaver W. Translation. In: Locke WN, Booth AD, editors. Machine Translation of Languages. Cambridge: MIT Press; 1955:15-23.
4. Dobrow MJ, Bytautas JP, Tharmalingam S, Hagens S. Interoperable electronic health records and health information exchanges: systematic review. JMIR Med Inform 2019 Jun 06;7(2):e12607 [FREE Full text] [doi: 10.2196/12607] [Medline: 31172961]
5. Deng H, Wang J, Liu X, Liu B, Lei J. Evaluating the outcomes of medical informatics development as a discipline in China: a publication perspective. Comput Methods Programs Biomed 2018 Oct;164:75-85. [doi: 10.1016/j.cmpb.2018.07.001] [Medline: 30195433]
6. Li Y, Hu J. Health informationization of China: status and development. Chin J Health Inform Manag 2012:001.
7. Gillum RF. From papyrus to the electronic tablet: a brief history of the clinical medical record with lessons for the digital age. Am J Med 2013 Oct;126(10):853-857. [doi: 10.1016/j.amjmed.2013.03.024] [Medline: 24054954]
8. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. Yearb Med Inform 2017 Aug;26(1):214-227 [FREE Full text] [doi: 10.15265/IY-2017-029] [Medline: 29063568]
9. Jung KY, Kim T, Jung J, Lee J, Choi JS, Mira K, et al. The effectiveness of near-field communication integrated with a mobile electronic medical record system: emergency department simulation study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11187 [FREE Full text] [doi: 10.2196/11187] [Medline: 30249577]
10. Bousquet C, Dahamna B, Guillemin-Lanne S, Darmoni SJ, Faviez C, Huot C, et al. The adverse drug reactions from patient reports in social media project: five major challenges to overcome to operationalize analysis and efficiently support pharmacovigilance process. JMIR Res Protoc 2017 Sep 21;6(9):e179 [FREE Full text] [doi: 10.2196/resprot.6463] [Medline: 28935617]
11. Kusch MKP, Zien A, Hachenberg C, Haefeli WE, Seidling HM. Information on adverse drug reactions-proof of principle for a structured database that allows customization of drug information. Int J Med Inform 2019 Sep 16;133:103970. [doi: 10.1016/j.ijmedinf.2019.103970] [Medline: 31704490]
12. Li F, Liu W, Yu H. Extraction of information related to adverse drug events from electronic health record notes: design of an end-to-end model based on deep learning. JMIR Med Inform 2018 Nov 26;6(4):e12159 [FREE Full text] [doi: 10.2196/12159] [Medline: 30478023]
13. Guo H, Na X, Hou L, Li J. Classifying Chinese questions related to health care posted by consumers via the internet. J Med Internet Res 2017 Jun 20;19(6):e220 [FREE Full text] [doi: 10.2196/jmir.7156] [Medline: 28634156]
14. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics 2016;36(1):176-191 [FREE Full text] [doi: 10.1148/rg.2016150080] [Medline: 26761536]
15. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016 May;279(2):329-343. [doi: 10.1148/radiol.16142770] [Medline: 27089187]
16. Névéol A, Zweigenbaum P. Clinical natural language processing in 2015: leveraging the variety of texts of clinical interest. Yearb Med Inform 2016 Nov 10(1):234-239 [FREE Full text] [doi: 10.15265/IY-2016-049] [Medline: 27830256]
17. Cobo M, Martínez M, Gutiérrez-Salcedo M, Fujita H, Herrera-Viedma E. 25 years at knowledge-based systems: a bibliometric analysis. Knowl-Based Syst 2015 May;80:3-13. [doi: 10.1016/j.knosys.2014.12.035]
18. Cobo M, López-Herrera A, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. J Informetrics 2011 Jan;5(1):146-166. [doi: 10.1016/j.joi.2010.10.002]
19. Chen X, Chen B. Discovering the recent research in natural language processing field based on a statistical approach. Lect Notes Comput Sci 2017;10676:507-517. [doi: 10.1007/978-3-319-71084-6_60]
20. Wallace ML, Larivière V, Gingras Y. A small world of citations? The influence of collaboration networks on citation practices. PLoS One 2012;7(3):e33339 [FREE Full text] [doi: 10.1371/journal.pone.0033339] [Medline: 22413016]
21. Chen X, Weng H. A data-driven approach for discovering the recent research status of diabetes in China. Lect Notes Comput Sci 2017;10594:89-101. [doi: 10.1007/978-3-319-69182-4_10]
22. Boudry C, Mouriaux F. Eye neoplasms research: a bibliometric analysis from 1966 to 2012. Eur J Ophthalmol 2015;25(4):357-365. [doi: 10.5301/ejo.5000556] [Medline: 25612654]
23. Chen X, Xie H, Wang FL, Liu Z, Xu J, Hao T. A bibliometric analysis of natural language processing in medical research. BMC Med Inform Decis Mak 2018 Mar 22;18(Suppl 1):14 [FREE Full text] [doi: 10.1186/s12911-018-0594-x] [Medline: 29589569]
24. Névéol A, Zweigenbaum P. Clinical natural language processing in 2014: foundational methods supporting efficient healthcare. Yearb Med Inform 2015 Aug 13;10(1):194-198 [FREE Full text] [doi: 10.15265/IY-2015-035] [Medline: 26293868]
25. Névéol A, Zweigenbaum P. Clinical natural language processing in 2015: leveraging the variety of texts of clinical interest. Yearb Med Inform 2016 Nov 10(1):234-239 [FREE Full text] [doi: 10.15265/IY-2016-049] [Medline: 27830256]

26. Névéol A, Zweigenbaum P. Making sense of big textual data for health care: findings from the section on clinical natural language processing. Yearb Med Inform 2017 Aug;26(1):228-234 [FREE Full text] [doi: 10.15265/IY-2017-027] [Medline: 29063569]
27. Li L, Zhang P, Zheng T, Zhang H, Jiang Z, Huang D. Integrating semantic information into multiple kernels for protein-protein interaction extraction from biomedical literatures. PLoS One 2014;9(3):e91898 [FREE Full text] [doi: 10.1371/journal.pone.0091898] [Medline: 24622773]
28. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097]
29. Van Eck NJ, Waltman L. How to normalize cooccurrence data? An analysis of some well-known similarity measures. J Am Soc Inf Sci 2009 Aug;60(8):1635-1651. [doi: 10.1002/asi.21075]
30. Li T, Ho Y, Li C. Bibliometric analysis on global Parkinson's disease research trends during 1991-2006. Neurosci Lett 2008 Aug 29;441(3):248-252. [doi: 10.1016/j.neulet.2008.06.044] [Medline: 18582532]
31. Gong Z, Xu G. Artificial Intelligence and its Applications. Beijing: Tsinghua University Press; 2015:343-345.
32. China mental health survey results summit. 2019 Apr 22. URL: http://news.china.com.cn/2019-04/22/content_74708259.htm [accessed 2020-01-02]
33. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer V, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
34. Perera G, Broadbent M, Callard F, Chang C, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. BMJ Open 2016 Mar 01;6(3):e008721 [FREE Full text] [doi: 10.1136/bmjopen-2015-008721] [Medline: 26932138]
35. Fidler MM, Gupta S, Soerjomataram I, Ferlay J, Steliarova-Foucher E, Bray F. Cancer incidence and mortality among young adults aged 20-39 years worldwide in 2012: a population-based study. Lancet Oncol 2017 Dec;18(12):1579-1589 [FREE Full text] [doi: 10.1016/S1470-2045(17)30677-0] [Medline: 29111259]
36. Zeng Z, Espino S, Roy A, Li X, Khan SA, Clare SE, et al. Using natural language processing and machine learning to identify breast cancer local recurrence. BMC Bioinformatics 2018 Dec 28;19(Suppl 17):498 [FREE Full text] [doi: 10.1186/s12859-018-2466-x] [Medline: 30591037]
37. Jones BE, South BR, Shao Y, Lu CC, Leng J, Sauer BC, et al. Development and validation of a natural language processing tool to identify patients treated for pneumonia across va emergency departments. Appl Clin Inform 2018 Jan;9(1):122-128 [FREE Full text] [doi: 10.1055/s-0038-1626725] [Medline: 29466818]
38. Zhang X, Tang W, Chen G, Ren F, Liang H, Dang Y, et al. An encapsulation of gene signatures for hepatocellular carcinoma, MicroRNA-132 predicted target genes and the corresponding overlaps. PLoS One 2016;11(7):e0159498 [FREE Full text] [doi: 10.1371/journal.pone.0159498] [Medline: 27467251]
39. Zhang R, Zhang L, Yang M, Huang L, Chen G, Feng Z. Potential role of microRNA-223-3p in the tumorigenesis of hepatocellular carcinoma: a comprehensive study based on data mining and bioinformatics. Mol Med Rep 2018 Feb;17(2):2211-2228. [doi: 10.3892/mmr.2017.8167] [Medline: 29207133]
40. Guo X, He T. Survey about research on information extraction. Comput Sci 2015:02.
41. Lei J, Tang B, Lu X, Gao K, Jiang M, Xu H. A comprehensive study of named entity recognition in Chinese clinical text. J Am Med Inform Assoc 2014;21(5):808-814 [FREE Full text] [doi: 10.1136/amiajnl-2013-002381] [Medline: 24347408]
42. Urbain J. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. J Biomed Inform 2015 Dec;58 Suppl:S143-S149 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.009] [Medline: 26305514]
43. Han J, Chen K, Fang L, Zhang S, Wang F, Ma H, et al. Improving the efficacy of the data entry process for clinical research with a natural language processing-driven medical information extraction system: quantitative field research. JMIR Med Inform 2019 Jul 16;7(3):e13331 [FREE Full text] [doi: 10.2196/13331] [Medline: 31313661]
44. Wang SY, Pershing S, Tran E, Hernandez-Boussard T. Automated extraction of ophthalmic surgery outcomes from the electronic health record. Int J Med Inform 2019 Oct 17;133:104007. [doi: 10.1016/j.ijmedinf.2019.104007] [Medline: 31706228]
45. Ware H, Mullett CJ, Jagannathan V, El-Rawas O. Machine learning-based coreference resolution of concepts in clinical documents. J Am Med Inform Assoc 2012;19(5):883-887 [FREE Full text] [doi: 10.1136/amiajnl-2011-000774] [Medline: 22582205]
46. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: analysis of chronic diseases with extracted entities and their relations. J Med Internet Res 2019 Jun 13;21(6):e12876 [FREE Full text] [doi: 10.2196/12876] [Medline: 31199327]
47. Doing-Harris K, Livnat Y, Meystre S. Automated concept and relationship extraction for the semi-automated ontology management (SEAM) system. J Biomed Semantics 2015;6:15 [FREE Full text] [doi: 10.1186/s13326-015-0011-7] [Medline: 25874077]

48. Karystianis G, Adily A, Schofield P, Knight L, Galdon C, Greenberg D, et al. Correction: automatic extraction of mental health disorders from domestic violence police narratives: text mining study. J Med Internet Res 2019 Apr 05;21(4):e13007 [FREE Full text] [doi: 10.2196/13007] [Medline: 30951492]
49. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
50. Usama M, Ahmad B, Xiao W, Hossain MS, Muhammad G. Self-attention based recurrent convolutional neural network for disease prediction using healthcare data. Comput Methods Programs Biomed 2019 Nov 11:105191. [doi: 10.1016/j.cmpb.2019.105191] [Medline: 31753591]
51. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using twitter data. J Med Internet Res 2017 Sep 12;19(9):e315 [FREE Full text] [doi: 10.2196/jmir.7393] [Medline: 28899847]
52. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]
53. Albalawi Y, Nikolov NS, Buckley J. Trustworthy health-related tweets on social media in Saudi Arabia: tweet metadata analysis. J Med Internet Res 2019 Oct 08;21(10):e14731 [FREE Full text] [doi: 10.2196/14731] [Medline: 31596242]
54. Garcia-Rudolph A, Laxe S, Saurí J, Bernabeu Guitart M. Stroke survivors on twitter: sentiment and topic analysis from a gender perspective. J Med Internet Res 2019 Aug 26;21(8):e14077 [FREE Full text] [doi: 10.2196/14077] [Medline: 31452514]
55. Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. J Med Internet Res 2019 Jun 27;21(6):e14199 [FREE Full text] [doi: 10.2196/14199] [Medline: 31250832]
56. Morioka C, Meng F, Taira R, Sayre J, Zimmerman P, Ishimitsu D, et al. Automatic classification of ultrasound screening examinations of the abdominal aorta. J Digit Imaging 2016 Dec;29(6):742-748 [FREE Full text] [doi: 10.1007/s10278-016-9889-6] [Medline: 27400914]
57. Amorim P, Moraes T, Fazanaro D, Silva J, Pedrini H. Shearlet and contourlet transforms for analysis of electrocardiogram signals. Comput Methods Programs Biomed 2018 Jul;161:125-132. [doi: 10.1016/j.cmpb.2018.04.021] [Medline: 29852955]
58. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019 Dec;132:103971. [doi: 10.1016/j.ijmedinf.2019.103971] [Medline: 31630063]
59. Kloehn N, Leroy G, Kauchak D, Gu Y, Colina S, Yuan NP, et al. Improving consumer understanding of medical text: development and validation of a new subsimplify algorithm to automatically generate term explanations in English and Spanish. J Med Internet Res 2018 Aug 02;20(8):e10779 [FREE Full text] [doi: 10.2196/10779] [Medline: 30072361]
60. Tang GY, Ni Y, Xie GT, Fan XL, Shi YL. A deep learning-based method for similar patient question retrieval in Chinese. Stud Health Technol Inform 2017;245:604-608. [Medline: 29295167]
61. Merabti T, Soualmia LF, Grosjean J, Palombi O, Müller J, Darmoni SJ. Translating the foundational model of anatomy into French using knowledge-based and lexical methods. BMC Med Inform Decis Mak 2011 Oct 26;11:65 [FREE Full text] [doi: 10.1186/1472-6947-11-65] [Medline: 22029629]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by E Borycki, G Eysenbach; submitted 28.10.19; peer-reviewed by K Chen, C Lovis, C Shivade, N Sundar Rajan; comments to author 19.11.19; revised version received 05.12.19; accepted 15.12.19; published 23.01.20.

Please cite as:
Wang J, Deng H, Liu B, Hu A, Liang J, Fan L, Zheng X, Wang T, Lei J
Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed
J Med Internet Res 2020;22(1):e16816
URL: http://www.jmir.org/2020/1/e16816/
doi:10.2196/16816
PMID:32012074

Viewpoint

Examining the Potential of Blockchain Technology to Meet the Needs of 21st-Century Japanese Health Care: Viewpoint on Use Cases and Policy

Tim Mackey[1,2,3*], MAS, PhD; Hirofumi Bekki[4*], MD, PhD; Tokio Matsuzaki[5], MD, PhD; Hiroshi Mizushima[6,7], PhD

[1]Department of Anesthesiology and Division of Global Public Health, University of California San Diego School of Medicine, La Jolla, CA, United States

[2]Global Health Policy Institute, La Jolla, CA, United States

[3]BlockLAB, San Diego Super Computer Center, La Jolla, CA, United States

[4]Department of Orthopedic Surgery, Kyushu University, Fukuoka, Japan

[5]Japan Biodesign Program, Department of Cardiac Surgery, The University of Tokyo, Tokyo, Japan

[6]Center for Public Health Informatics, National Institute of Public Health, Saitama, Japan

[7]Japanese Medical Blockchain Association, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Tim Mackey, MAS, PhD
Department of Anesthesiology and Division of Global Public Health
University of California San Diego School of Medicine
8950 Villa La Jolla Drive
A124
La Jolla, CA, 92037
United States
Phone: 1 9514914161
Email: tmackey@ucsd.edu

Abstract

Japan is undergoing a major population health transition as its society ages, and it continues to experience low birth rates. An aging Japan will bring new challenges to its public health system, highlighted as a model for universal health coverage (UHC) around the world. Specific challenges Japan's health care system will face include an increase in national public health expenditures, higher demand for health care services, acute need for elder and long-term care, shortage of health care workers, and disparities between health care access in rural versus urban areas. Blockchain technology has the potential to address some of these challenges, but only if a health blockchain is conceptualized, designed, localized, and deployed in a way that is compatible with Japan's centralized UHC-centric public health system. Blockchain solutions must also be adaptive to opportunities and barriers unique to Japan's national health and innovation policy, including its regulatory sandbox system, while also seeking to learn from blockchain adoption in the private sector and in other countries. This viewpoint outlines the major opportunities and potential challenges to blockchain adoption for the future of Japan's health care.

(J Med Internet Res 2020;22(1):e13649) doi:10.2196/13649

KEYWORDS

information storage and retrieval; blockchain; Japan; aging; health informatics; health policy; global health

Introduction

Background

Blockchain technology is increasingly gaining interest around the world and is poised to be a technology that rapidly globalizes, now over a decade since the author under the Japanese name Satoshi Nakamoto first published the bitcoin white paper that led to popularization of the technology. Fundamentally, blockchain is a form of distributed digital ledger technology used to share and store data in a decentralized manner [1]. *Blocks* of data are secured through cryptography so that a *chain* of blocks is created that provides the provenance of any given transaction and also makes the records tamper evident [2]. In the context of health care, a *health* blockchain can enable better trust, security, management, and transparency

of health care data, processes, and transactions and is actively being explored as a potential tool to improve the delivery of health care in several countries [3].

Despite growing attention, investment, and ongoing efforts toward commercialization, blockchain technology is still in its early adoption phases, particularly when comparing the health care industry to other sectors such as cryptocurrency markets, financial technology, and blockchain for logistics [4]. Importantly, blockchain solutions need to be tailored to the challenges unique to health care, including issues related to patient and data privacy, coordinating care across multiple actors (including hospital, payers, health care providers, pharmaceutical and device manufacturers, health technology providers, and patients), ensuring appropriate sharing, and maintaining integrity of health care data, while also being responsive to regulatory complexity [3,5].

With its potential to transform different industry verticals, several proofs of concept, pilot projects, and solutions in transition between development and production are being explored by the private sector (including large to midsize companies and startups), consortiums, and national governments who are specifically focused on health care [3,6,7]. Beyond tailoring blockchain solution design to the health care sector, blockchain use cases also need to be localized to the specific community and population health needs faced by an individual country and also be compatible with its national health system design.

Localizing Blockchain for Health and Japan

The need to localize blockchain solutions includes countries such as Japan, the third largest economy in the world (by nominal gross domestic product [GDP]), which has recently seen increased activity around assessing blockchain applications in the health care sector that are specific to its unique and emerging clinical and public health challenges [8]. Specifically, Japan operates a robust public health system, characterized by universal health coverage (UHC), which has long been touted by the Japanese government as first class in terms of access, quality, and low cost [9].

Despite Japan's strong public health system and commitment to UHC, its 21st-century health care challenges are historically unique. Central to these challenges is a demographic shift that is underway called *koreikashakai* (translated as *population aging society*), which has resulted in Japan becoming the world's oldest country (with 27.7% [n=35.15 million] of its population aged older than 65 years) [10]. The resulting *upside down* demographic pyramid (precipitated by low birth rate and high life expectancy) will bring new economic challenges (eg, stabilizing funding of national pensions with concomitant need for growing availability of public services) coupled with increased strain on national health care capacity. Specific challenges include increased national public health expenditures, higher demand for health care services lacking appropriate cost controls, acute need for elder and long-term care, lack of availability and shortage of health care workers (including nurses and caregivers), and continuing disparities between health care access in rural versus urban areas [11-13].

Some of these challenges are becoming more acute, with recent data indicating that medical-related spending is growing in Japan [14]. Furthermore, although Japan operates a national health insurance system and a centralized social security and tax number system (known as *My Number system*), adoption of centralized health informatics infrastructure is lagging. This includes electronic medical records, where, in 2017, the adoption rate in Japan was reported as only 34.4% according to the Japan Country Commercial Guide published by export.gov, although implementation rates are projected to increase [15]. Hence, the digitization of Japan's health system, which could enable creation of centralized health care data management and decision making remains limited, despite an acute need to lower health care–related administrative costs and provide more efficient access for increasing demand of health care services.

All these factors give rise to promising prospects for blockchain solutions to address specific needs of Japan's health care system, but only if they are conceptualized, designed, and deployed to align with the country's national health identity that focuses on a UHC-centric public health system. In response, this viewpoint outlines the unique opportunities and challenges faced by Japan's current and future health care system, examines how these challenges can be addressed by blockchain technology, and also discusses emerging Japanese public policy on technology and whether it can encourage blockchain adoption.

Japan's Current and Future Health Care Challenges

First, it is important to understand the characteristics of Japan's health care system and the current challenges it faces today and in the future. Japan's health care system is based on a centralized UHC public insurance system with care provisioned by a network of more than 4000 public and private payers [16]. Residents of Japan are required to have health insurance coverage and receive coverage through Employees' Health Insurance (*Kenko-Hoken*) or the national health insurance system (*Kokumin-Kenko-Hoken*). Citizens without insurance coverage from employers can participate in the national health insurance program. Depending on the total income and age of the insured, the ratio of medical fees patients pay differs from 0% to 30%, with the government paying the remaining fees [17]. The national health insurance system is based on fiscal resources generated by a combination of employee and employer contributions, cost sharing by patients, and subsidization by the government, all factors that are impacted by demographic changes.

The impact of demographic changes on public health care access, provisioning, and financing is central to technology development. Japan is now a rapidly aging society with the number of elderly people (aged older than 65 years) quadrupling in the last 40 years with some projections also suggesting that this will lead to a long period of overall population decline [18]. In fact, the number of elderly is expected to peak at a staggering 39.35 million by 2042, which will represent more than one in three people being elderly [10,19]. Conversely, the population of the labor force (aged 15-64 years) has been decreasing since 1995 when 87 million people were included in this group but,

in 2016, only comprised 76 million people [10]. The labor force is further estimated to shrink to 68 million in 2030, also coinciding with an increase in nonworking pensioners [10]. This will lead to a decrease in working age individuals and employers who can contribute payroll taxes to the health system to fund public health programs [10]. It will also shift the risk pool of enrollees to more expensive patients who require higher frequency visits, long-term care, and more complex health interventions.

All these demographic trends point to a trifecta of health system shocks, including increased utilization, rising national medical costs, and decreases in health care financing relative to population and pensioner changes [10]. In 2015, Japan's total health spending accounted for 11.2% of its GDP equating to 42.3 trillion yen and is now ranked third in total health spending out of 35 Organisation for Economic Co-operation and Development (OECD) countries [14,20]. Japan's increase in health spending of GDP also reflects an overall trend of steadily increasing global expenditures on health care, including a projected 9% of GDP allocated globally to health spending by 2040 [21]. Japan's elderly care health expenditure is also projected to rapidly increase, estimated to peak in the next few decades and then continue increasing until 2065 [22]. Given these characteristics of the *aging of Japan* and its public health system design, it is expected that increased costs related to a growing burden of chronic diseases and associated high-cost medical interventions and technologies will result in a health care funding gap of approximately 44 trillion yen by 2035 [16].

These factors are also exacerbated by suboptimal health care utilization in the country. For example, Japanese patients tend to go to outpatient clinics more often than in other OECD countries; Japanese physicians see approximately twice as many patients annually compared with other countries, and the length of hospital stay is very high [14]. The provisioning of unnecessary health care services, which includes higher volume or higher costs, is also something Japan is struggling to tackle. In addition, Japan aggressively introduces and uses advanced medical devices and new health technology, which increase the cost of diagnosis and disease management [23].

Expensive medical services include radiographic examination procedures, such as computed tomography and magnetic resonance imaging, which result in Japanese patients having higher exposure to radiation compared with other developed countries and concomitantly higher costs because of the frequency of these procedures [24]. Overutilization is incentivized by a fee-for-service model coupled with a national pricing structure that is meant to control costs but not utlization [25]. Overutilization may also be related to lack of comprehensive facility accreditation, as, currently, the Japan Council for Quality Health Care (JCQHC; established in 1995) acts as third-party accreditor to evaluate the functions of medical institutions, but only 26.2% (n=2192) of hospitals nationwide are certified and reviewed by the JCQHC [26].

In addition, with the increasing number of elderly patients, more physicians and other health care professionals (including nurses and caregivers) will be needed in the health care workforce [11,12]. According to a survey by the National Institute of Population and Social Security Research in Japan, the rate of total number of medical doctors was 2.4 per 1000 people in 2010, which is fewer than the average of other OECD countries [11]. Physician shortages are also impacted by a system where physicians can decide their specialties freely regardless of grades or achievements. In terms of board certification, historically, Japan has not set a limitation on the number of doctors in each department that approves these specialties leading to specialty imbalances. In addition to physicians, the Ministry of Health, Labour and Welfare (MHLW) has also reported a gap of 2 million nursing personnel as would be required by 2025 to meet health care utilization demands [12]. Furthermore, Japan's immigration policies and bilateral trade agreements have also negatively impacted availability to foreign nursing and caregiver personnel [27,28].

Finally, there exists a growing gap of medical care coverage between rural and urban communities. In 2017, the reporting agency Nikkei, Inc, conducted an analysis on government data and reported that the mortality rate from acute myocardial infarction differed 4 to 5 times depending on whether a patient resided in a rural or urban community [13]. These data suggest that the quality of medical care may experience variation depending on geographic location (including rural vs urban) and that these differences can manifest in different communities even in the same prefecture.

In summary, Japan's current and future health care challenges are largely driven by its rapidly aging demographic and how it impacts national health care expenditure, utilization, and workforce demands in a system that champions UHC. Importantly, other countries are also experiencing some of these challenges including aging populations and lower birth rates, higher burden of chronic diseases (including lower income countries experiencing an epidemiological shift from communicable to noncommunicable diseases), and lack of access to rural health care, but perhaps not to the same extent Japan is facing because of its rapidly changing yet still largely homogenous demographics (in 2018, Japan hit a record of 2.497 million foreigners, but this still represents only 1.99% of the entire population) [29].

Although technology is not the panacea for all these problems, blockchain in conjunction with other digital health solutions has the potential to lower health care costs, enable extension and broader access to health care services, combat health care fraud, and increase the efficiency of the health care workforce. We explore some existing and potential health blockchain use cases for Japan below.

Availability of Data and Materials

Data associated with our multilingual literature and legal, and policy review are available via information in references.

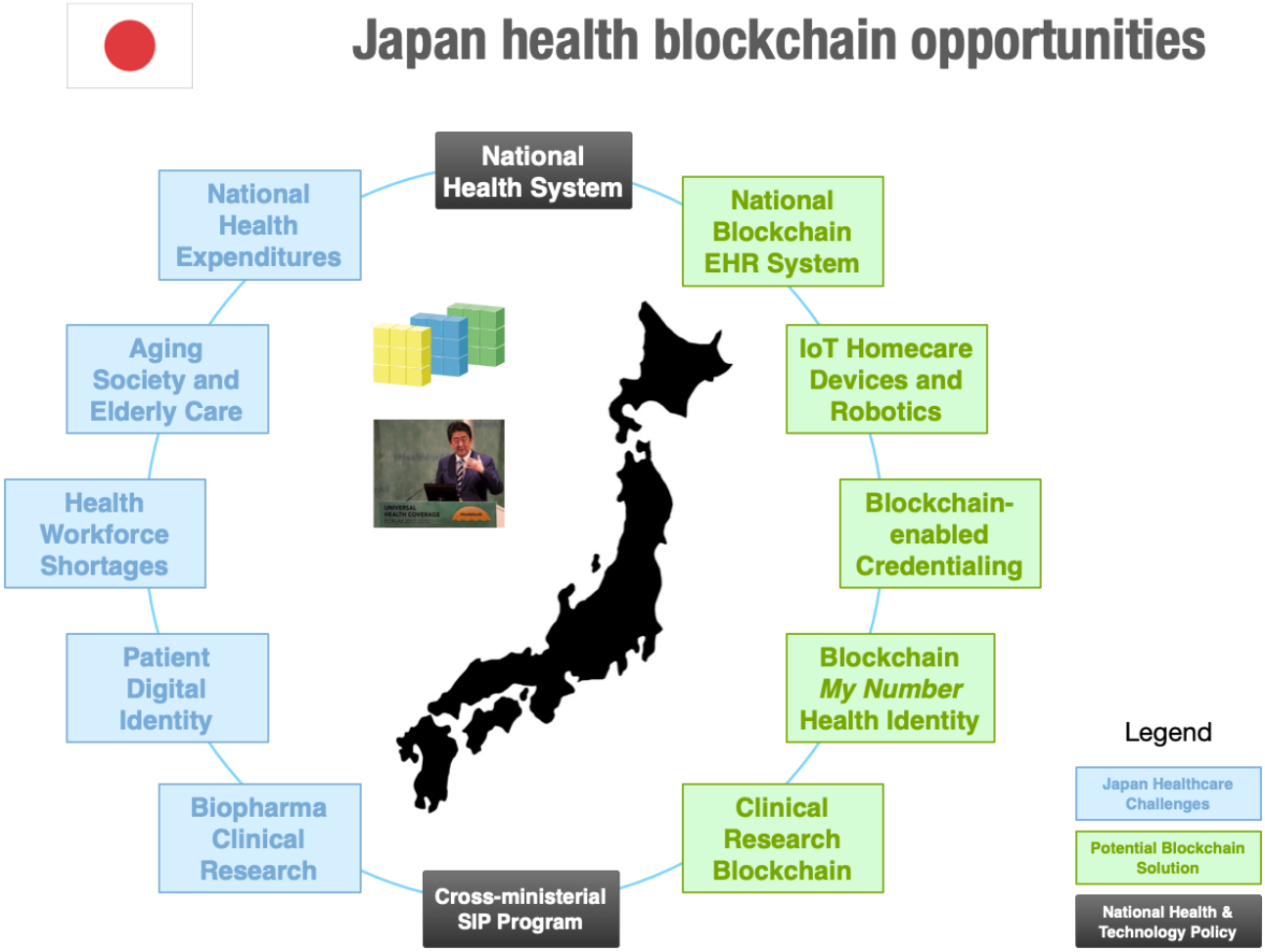
Blockchain: A Potential Solution for Japan's Future Health Care Landscape?

Given the unique current and future challenges faced by Japan's national health system, what challenge areas are blockchain

technology solutions uniquely positioned to offer real-world solutions? The answer to that question resides in mapping use cases to Japan's UHC-based health care system, examining how they could be localized, and also taking into account the national policy and health care information technology (IT) architecture of a centralized public health system (see Figure 1 for a summary of thematic blockchain opportunities).

Current and future adoption of blockchain solutions in Japan may also originate from use cases that do not directly address national health system priorities. Instead, these use cases might target Japan's private sector and commercial needs, particularly relevant as Japan is the world's third largest pharmaceutical market and one of the world's largest markets for medical devices [30,31]. We explore select Japanese private and public sector blockchain health care use cases in the following sections, with some examples of use cases which are summarized in Multimedia Appendix 1.

Figure 1. Summary of blockchain potential for the Japanese health care system. EHR: electronic health record; IoT: Internet of Things; SIP: Cross-ministerial Strategic Innovation Promotion Program.



Private Sector Blockchain Use Cases

Translating blockchain solutions to the Japanese system context can first start by examining many of the leading use cases for health and life science blockchain deployments in other markets. These include management of electronic health records (EHRs), optimizing the performance of clinical trials and enhancing the integrity of the health and pharmaceutical supply chain [3,32-36]. Many of these use cases are more general to improving the competitiveness of Japan's biotechnology, pharmaceutical, medical device, and life sciences private sector industries and not specific to its national health care system but still merit examination and could drive public sector adoption.

A leading health care blockchain area of development is clinical research blockchains that can produce verifiable data associated with patient recruitment, enhance sharing of clinical research data with patients and clinical trial study protocol data management, and, when provided a set of core defined metadata, can help ensure clinical trial integrity, transparency, and auditability to regulators [37,38]. Clinical trial blockchain solutions and consortiums are increasingly becoming more active, with major pharmaceutical manufacturers such as Boehringer Ingelheim, Pfizer, Amgen, Sanofi, and Bayer announcing initiatives in this area [3,39].

The Japanese startup company, Susmed Inc, recently announced a pilot project that has been certified by MHLW and the Ministry of Economy, Trade, and Industry for securing reliable clinical data monitoring using blockchain technology and mobile health (mHealth; mobile phone) apps [40,41]. In addition, the Japanese pharmaceutical manufacturer, Takeda, announced it was co-chairing with Microsoft and ERORDIS-Rare Disease Europe the *Global Commission to End the Diagnostic Odyssey for*

Children aimed at addressing barriers to developing diagnostics for patients with rare diseases, which includes a pilot project examining the use of blockchain to develop secure patient registries and a rare disease patient passport [42]. A clinical research blockchain system could have also helped to avoid the 2014 arrest of a former Japanese employee of Novartis who was accused of falsifying clinical data for the popular hypertension drug, valsartan [43].

Similarly, blockchain solutions that are being developed to better secure the integrity of pharmaceutical supply chain in the United States and the European Union may also have utility in addressing infiltration of substandard and falsified medicines in Japan's health system and also assist Japanese pharmaceutical manufacturers in securing international distribution of their own products [3]. Several companies are currently developing blockchain solutions for the pharmaceutical supply chain, including MediLedger, IBM, and SAP. They are primarily examining how blockchain can be used to establish provenance of drug pedigree and how it can enable regulatory adherence to national track-and-trace requirements, and they have also examined the potential utility of combating infiltration of substandard and falsified drugs [3,44].

In fact, there have been reports of counterfeit versions of erectile dysfunction drugs and weight loss pills being imported via internet sales and even counterfeit hepatitis C medication being detected in legitimate Japanese pharmacy chains [45-47]. Furthermore, the MHLW has established a *Suspicious Drugs Reporting Network* demonstrating growing public concern over the issue, particularly in the context of illegal internet pharmacies that operate outside the legitimate supply chain [48]. In addition, unrelated to the pharmaceutical supply chain but relevant to supply chain integrity, IBM and Asahi Refining (part of Japanese company Asahi Holdings) are part of a consortium called the TrustChain Initiative that will develop a blockchain network for establishing provenance for jewelry tracing that could show utility for health care commodities [49].

Blockchain Solutions for Japan's National Health System Challenges

Beyond commercially viable blockchain solutions for Japan's private sector, some blockchain use cases map directly to Japan's future demographically led health care challenges as previously summarized. These use cases include (1) blockchain-based EHR systems to improve health care billing, utilization, and reducing waste to ensure efficiency in national health care administration and financing; (2) blockchain-enabled provider and patient directories while also unifying services together under a blockchain digital identity tied to national social security system records; and (3) blockchain integration with Internet of Medical Things (IoMT) to address the aging society by enabling secure and verifiable home care delivery and telehealth.

Blockchain-Based Medical Record Systems

First, although Japan is a public health system, electronic records are generally managed by individual institutions and are housed locally on disparate systems operated by private firms, not in a centralized database by the national government. Hence, patients' data and health identity are scattered across different providers, which impedes access and portability, while also limiting big data analysis for insights into population health trends. The Japanese government also currently prohibits mixed billing, which consists of private and public health insurance for 1 condition; therefore, patients often have no choice but to select the treatment covered by their insurance [50]. This policy decision means future EHR and reimbursement blockchains will need to be responsive to centralized financing, although health care systems and providers themselves may be decentralized in operating and managing their health IT systems and data. For example, even radiographic examinations performed in Japan on the same day can be conducted at different provider locations, with the images residing in different databases. Differences in health system design and policy means that there will be multiple trade-offs in the design, development, and implementation of blockchain EHR systems related to utility of the system, security, and scalability that could be informed by the Model National Health Service blockchain as proposed by O'Donoghue et al [51].

Importantly, lagging adoption of health information exchange (HIE) despite a system that is centralized and is single payer provides an opportunity for blockchain to act as an intermediary or locator service to make queries of traditional database-bound medical records from different providers. Although there are signs that HIE between different health care systems in Japan is spreading, a privacy framework and mandated EHR data sharing across all 47 prefectures remains absent [52-54]. This problem seems well situated for blockchain adoption, including exploring managing patient digital data and identity through a shared distributed ledger that can be accessed by providers and is tied to the *My Number* social security and national tax number ID system. Despite the potential promise of blockchain-enabled HIE, concerns about patient privacy, coordination, opt-in to applications, and specific information exchange scenarios need to be further studied [55].

Examples that Japan could examine to assess the initial viability of a national health blockchain EHR system include a recent announcement by Taipei Medical University of the launch of a blockchain solution to improve patient referral services and integrate health care networks to enable better access to medical records [56]. The system includes participation from more than 100 community-based clinics, uses smart contracts, and enables access to EHR data [56,57]. Importantly, Taiwan also operates a national single-payer publicly funded health system under its National Health Insurance that covers 99% of the population and involves both public and private providers [58]. In this sense, it operates a health care system with many similarities to Japan, including challenges related to health data management and sharing, and hence could be informative to Japan's own blockchain development around EHRs [58].

Other countries, such as Estonia that has adopted widescale use of electronic health (eHealth) approaches to its national health care system (including a Web-based eHealth record, electronic ID card, and electronic patient portal), are also assessing blockchain for purposes of maintaining security and integrity of health records and could also be informative [59,60]. There are also several companies, initiatives, consortiums, and research

groups looking at blockchain EHR integration in the United States. This includes MedRec, an open-source platform that uses blockchain and smart contracts to create a record of patient-provider interactions and access and viewing permissions of medical records [61,62].

Blockchain Directories and Unification Under My Number System

Another concrete example of blockchain use in health care data is blockchain-based provider data management and directory systems that attempt to reconcile health care provider credentialing and national medical licenses, which can be subject to fraud and error, including cases of identity fraud involving those pretending to be physicians in Japan [3,63,64]. Hence, a blockchain provider directory can provide a single source of verified data that can be better shared across different hospitals and payers, such as the one currently being explored by the Synaptic Health Alliance (which includes large US health care organizations such as Aetna, UnitedHealth, and Humana) [65,66]. Such a system could also be integrated with existing efforts by the Japan Medical Association to digitize medical credentialing under its JMA Electronic Certification Center using smart cards, electronic signatures, and identity authentication [64].

Compared with the complexity of the US health care system, which includes both public and private payers and providers, instituting a provider blockchain directory in Japan's single-payer UHC structure should be a more straightforward task. A blockchain-based provider directory also opens the door for the development of a public national patient-centered directory with health care access verified by the *My Number* system. This could provide patients and providers with EHR verification (tied to validated *My Number* digital identity), recording and auditability of requests for data access, portability of health care data, enhance verification of health care claims, and potentially improve the continuity of care as pointers to EHRs could be linked to the *My Number* system across different Japanese health care providers [67].

Importantly, this design could adapt well to Japan's current decentralized network of health care IT management where EHRs may continue to reside within each individual health care organization with hashed pointers enabling sharing in a secure and distributed fashion while also encouraging patient-driven interoperability [62,68]. Access could be based on principles of patient-centered permissions under a *My Number* digital identity and/or corresponding digital wallet, effectively mitigating the possibility for social welfare or health care fraud (eg, recently a woman in her 70s was defrauded several million yen in a scam associated with her My Number system ID) [69]. Finally, health data could also be aggregated and deidentified for population health research purposes [67].

Integrating Blockchain to Japan's Internet of Things Ecosystem

Integrating blockchain into the Internet of Things (IoT) and more specifically IoMT to better enable home care and telehealth services will also be a leading future use case in Japan, especially given the rise of connected medical devices and the ubiquitous use of other tools that can enable medical applications (such as mobile phones). Specifically, Japan is a country with widespread mobile phone adoption with 92% of adults reporting they own a mobile device according to a survey by the Pew Research Center among 30,133 people in 27 countries in 2018 [70]. Hence, increased health care utilization in the inpatient setting because of growth in the number of elderly patients and continued shortages of health care workers presents opportunities for expansion of Japan's collective connected health offerings of telemedicine, IoMT, and robotics industry.

All these forms of connected health technology will be critical to increase efficiencies, lower costs, and address health care workforce capacity issues. Furthermore, lack of intrinsic security measures can mean that health data on IoT devices may be vulnerable to threats, such as relying on a single gateway for data to be breached or failure to validate access or secure data when exchanged [71]. Blockchain has the potential to address these challenges by enabling the patient to monitor and control access and security to their data (including through the use of private encryption keys or smart contracts) that would ensure a higher level of autonomy, privacy, and control [71,72]. From a data integrity perspective, once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, rendering data tamper evident, and better securing health care data for purposes of analysis, insights into patient compliance and treatment, and also potentially enhancing reimbursement processes [73].

Blockchain's potential potentially extends to Japan's growing landscape of telehealth expansion. Research indicates that the focus areas for telemedicine development in Japan include increasing access to health care for rural and remote communities, enabling telemedicine in home care, and use for prevention and lifestyle modification [74]. Hence, the potential for blockchain-enabled telemedicine services, such as those being piloted in London, UK by the company Medicalchain, has the potential to secure telemedicine data (including Web-based and video consultations), create linkage with EHR data with access requests being recorded on a distributed ledger, and also enable payment via cryptocurrencies [75]. Japan's potential for telehealth expansion coincides with recent deregulation of the telemedicine industry, where health insurance can now be used for telemedicine consultations [76]. The MHLW has also expressed its overall support of telemedicine in its proposal *Japan Vision: Health Care 2035* with Japanese health care startups actively providing remote consultation services for different conditions to Japanese health care institutions and their patients [76].

Additional opportunities for the mHealth and IoMT blockchain market are also emerging, as Japan represents one of the world's largest markets for medical devices (valued at $33.3 billion in 2013) and as it represents a major export market, largely driven by its aging population that relies on devices for health maintenance and treating quality of life and age-related conditions [31]. Blockchains can enable interconnection of smart devices to collect health care data (including in the acute care setting, in the outpatient setting, and for home care) while also verifying the identity and provenance of data that may originate from multiple IoT-enabled sources [41]. Demonstrating

early experimentation in this area of mHealth, Susmed Inc published a paper in 2017 in *JMIR mHealth and uHealth* describing the development of a blockchain-based tamper-resistant mHealth system using a smartphone app to deliver cognitive behavioral therapy for insomnia [41]. This in-country–driven innovation represents a unique combination of using blockchain as an mHealth intervention while also better securing data for use in clinical research [41].

Finally, Japan also has a robust robotics industry, including growing interest in the health care sector where robots are being tested for use in elderly care to address senior care workforce shortages, with projects enjoying support from the Japanese government [77-79]. In this context, blockchain has been proposed as a solution to better distribute and secure information for robotic swarm operations and make them more efficient, representing a potential frontier technology application for blockchain in health care [80,81].

Decentralized Implementation?

Although Japan's health care system has a centralized UHC structure, health blockchain implementation in Japan will likely be decentralized, meaning that it will not at first take an integrated approach. Although it may be ideal that all the blockchain use cases described above could be integrated into a single national blockchain technology framework and governance architecture, it is more likely that early implementation will occur through deployment of special-purpose blockchains specific to discrete health care challenges.

Our review finds that private and public sector blockchain use cases can have different goals and design elements (eg public vs private or consortium-based designs, differing permissions and privacy needs, varying forms of data integration and sharing, and may require different regulatory approaches). Hence, uniquely situated private sector versus public health sector needs may necessitate customized and adaptive policy making in combination with technical standard setting to better enable interoperability and shared benefit. Importantly, even in a centralized health system similar to Japan, one size will not fit all and Japan health blockchain deployments will need to be *fit-for-purpose* for specific health care challenges and the needs of their different local, provincial, and national-level stakeholders [3].

Conclusions

Although Japanese health care blockchain use cases hold promise, progress in research and development, financial investment, and eventual deployment will require national government buy-in (as Japan's health system is largely publicly funded) coupled with adaptive policy making to ensure blockchain technologies are incentivized and regulated correctly. Fortunately, the Japanese government is already making efforts to actively explore innovative and disruptive technologies through a regulatory sandbox system under the Cabinet Secretariat [82,83]. This system allows domestic and foreign organizations to apply for demonstration and evaluation of new technology, such as blockchain and IoT, without being subject to existing regulations while also opening up the possibility for future deregulation measures [82]. In fact, there are a few Japanese blockchain companies already taking advantage of the regulatory sandbox, including Susmed, Inc, which has 2 blockchain pilot projects for a clinical data management system [40].

Furthermore, leading Japan's efforts in new technology spaces is the Cross-ministerial Strategic Innovation Promotion Program (SIP). The SIP is a national project for science, technology, and innovation supported by Japan's Cabinet office. As an example of some of their efforts in emerging technology, the SIP is engaged in projects to establish artificial intelligence and big data solutions in hospitals to serve patients with automated diagnosis and treatment options through a commitment of $20 million in funding for 2018 [84]. One of the goals of the SIP is to establish a medical database with strong security, which is called subtheme A. Interestingly, according to the draft for subtheme A, blockchain technology will be explored as a potential solution for this project area. However, the Japanese government also hinted at some practical barriers including concerns that the costs of blockchain systems are still unknown and that there is a shortage of qualified blockchain developers. These real-world challenges have the potential to impede future government blockchain adoption in the health sector despite emerging policy support [85].

Collectively, government efforts and early commercial interest present an opportunity for shared decision making to develop technical standards around blockchain complemented by efforts of the SIP and regulatory sandbox regime. In fact, organizations such as the Institute of Electrical and Electronics Engineers (IEEE) Standards Association are working on a series of technical standards focused on distributed ledger technology use in the health care and life and social sciences (P2418.6), IoT (P2418.1), and applications in government (P.2418.8). Japan's own IEEE Japan Council and its affiliated sections could contribute to development of these distributed ledger technology standards that could also be localized for Japan's specific health, industry, and technology sector needs. This could also be coordinated through regulatory and policy coordination with the MHLW as the lead health agency and acting in conjunction with other nonhealth agencies working to encourage adoption of innovative technologies.

In addition, lessons for the future of Japan's health blockchain can also be learned from the maturation of the financial services technology (also known as fintech) sector in the country. Blockchain's earliest and most popularized use cases have arisen from cryptocurrencies, and more specifically bitcoin. Not surprisingly, Japan's most mature sector for blockchain adoption has been fintech with several ongoing initiatives by Japanese banks and financial institutions to create cryptocurrencies and decentralized digital currency marketplaces and exchanges [86,87]. For example, Mitsubishi UFJ Financial Group, the largest bank in Japan, is planning on launching a bitcoin and cryptocurrency exchange targeting institutional and retail investors. In addition, Oversea-Chinese Banking Corporation, a bank in Southeast Asia, has carried out successful pilots of payment transactions using blockchain technology [88].

Overall, these fintech blockchains are designed to show that traditional functions of the financial system can be reliably executed by decentralized networks and in doing so speed up financial settlements while also raising the prospect of new financial system design [89]. However, there have been calls for regulation and taxation of fintech-related activities including blockchain, which include changes to Japan's Financial Services Agency [90]. Hence, these early adoption challenges for fintech in Japan raise important questions regarding future adoption and regulation of blockchain technologies in other industries that are highly regulated and have privacy considerations such as health care.

Importantly, any Japan health blockchain strategy has to take into account the national and public health characteristics of Japan's health system in addition to the unique challenges it faces with its aging society, concerns of health care overutilization, and shortage in its health care workforce. One future approach may be to design Japan's health care blockchains as part of a Whole-of-Government Approach similar to Estonia where an electronic, national, government-backed blockchain is being used in the health care sector [91]. Regardless of the strategy, Japan's current health care problems are acute, and its future health care challenges cannot be solved by a single solution, as technology is only one potential component of addressing complex population health challenges. In this sense, blockchain technology can act as an important enabler for technology and governance solutions for 21st-century Japanese health care, but only if it is localized, fit-for-purpose, and meets the needs of the Japanese people.

Acknowledgments

This research was partly supported by the Japan Health, Labour and Welfare Sciences Research Grants for Research on Policy Planning and Evaluation. Authors would like to acknowledge and thank this support.

Authors' Contributions

HB and TM jointly collected the data, designed the study, conducted the data analyses, and wrote the manuscript. All authors contributed to the formulation, drafting, completion, and approval of the final manuscript.

Conflicts of Interest

TKM has received speaker fees and reimbursement for travel from Cardinal Health to present on blockchain and health care research. He is also the co-chair of the IEEE Standards Association Supply Chain/Clinical Trials Technology Implementation Industry Connections Program that focuses on stakeholder collaboration around blockchain technology for the pharmaceutical supply chain and has received reimbursement for travel expenses associated with speaking at IEEE-sponsored events. In addition, he is a noncompensated advisory board member for the blockchain and pharmaceutical company FarmaTrust. All other authors report no conflicts of interest associated with this paper.

Multimedia Appendix 1
Summary of Japan health blockchain use cases.
[DOCX File , 15 KB - jmir_v22i1e13649_app1.docx]

References

1. Nakamoto S. Bitcoin. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System URL: https://bitcoin.org/bitcoin.pdf [accessed 2019-10-07]
2. Harvard Business School. 2015. The Great Chain of Being Sure About Things URL: https://www.hbs.edu/faculty/conferences/2015-strategy-research/Documents/The great chain of being sure about things.pdf [accessed 2019-10-06]
3. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med 2019 Mar 27;17(1):68 [FREE Full text] [doi: 10.1186/s12916-019-1296-7] [Medline: 30914045]
4. Gartner. 2018 Oct 16. The Reality of Blockchain URL: https://www.gartner.com/smarterwithgartner/the-reality-of-blockchain/ [accessed 2019-10-07]
5. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 1;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
6. Innovation Enterprise. Blockchain Top Trends In 2017 URL: https://channels.theinnovationenterprise.com/articles/blockchain-top-trends-in-2017 [accessed 2019-10-07]
7. Agbo C, Mahmoud Q, Eklund J. Blockchain technology in healthcare: a systematic review. Healthcare (Basel) 2019 Apr 4;7(2):pii: E56 [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
8. Radanović I, Likić R. Opportunities for use of blockchain technology in medicine. Appl Health Econ Health Policy 2018 Oct;16(5):583-590. [doi: 10.1007/s40258-018-0412-8] [Medline: 30022440]

9. Otake T. The Japan Times. 2017 Feb 19. Japan's Buckling Health Care System at a Crossroads URL: https://www.japantimes.co.jp/news/2017/02/19/national/japans-buckling-health-care-system-crossroads/#.XBFPgqeZMws [accessed 2019-10-07]
10. Cabinet Office. 2018. URL: https://www8.cao.go.jp/kourei/whitepaper/w-2018/gaiyou/pdf/1s1s.pdf [accessed 2019-10-07]
11. National Institute of Social Security and Population Issues. Chapter 4 Health Care URL: http://www.ipss.go.jp/s-info/e/ssj2014/pdf/04_SSJ2014.pdf [accessed 2019-10-07]
12. Japanese Nursing Association. 2016. Nursing in Japan URL: http://www.nurse.or.jp/jna/english/pdf/nursing-in-japan2016.pdf [accessed 2019-10-07]
13. Nikkei Visual Data. 2018. Regional Disparities in Cancer, Heart Disease, and Stroke: What is Your City? URL: https://vdata.nikkei.com/newsgraphics/health-expenditures-topics1/ [accessed 2019-10-04]
14. OECD iLibrary. 2015. Health at a Glance 2015 URL: http://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2015_health_glance-2015-en [accessed 2019-10-02]
15. Export.gov - Helping Businesses Export. 2018. Japan - Healthcare IT URL: https://www.export.gov/article?id=Japan-healthcare-IT [accessed 2019-10-04]
16. McKinsey & Company. 2018. Improving Japan's Health Care System URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/improving-japans-health-care-system [accessed 2019-10-04]
17. Ministry of Health Labour and Welfare. An Outline of the Japanese Medical System URL: https://www.mhlw.go.jp/english/policy/health-medical/health-insurance/index.html [accessed 2019-10-05]
18. National Institute of Social Security and Population Issues. 2017 Jul 31. Population Projections for Japan:2016-2065 URL: http://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/pp29_ReportALL.pdf [accessed 2019-10-05]
19. National Institute of Population and Social Security Research. Population Projections for Japan: 2016 to 2065 URL: http://www.ipss.go.jp/pp-zenkoku/e/zenkoku_e2017/pp_zenkoku2017e.asp [accessed 2019-10-07]
20. Reuters. 2018 Jan 23. Japan's Fiscal Target Delay Fuels Worry About Healthcare Spending URL: https://www.reuters.com/article/japan-economy-budget-idUSL4N1PI20K [accessed 2019-10-05]
21. Institute for Health Metrics and Evaluation. 2016 Apr 13. Global Spending on Health is Expected to Increase to $18.28 Trillion Worldwide by 2040 but Many Countries Will Miss Important Health Benchmarks URL: http://www.healthdata.org/news-release/global-spending-health-expected-increase-1828-trillion-worldwide-2040-many-countries [accessed 2019-10-05]
22. Fukawa T. Elderly Population Projection and Their Health Expenditure Prospects in Japan. Mod Econ 2017;08(11):1258-1271 [FREE Full text] [doi: 10.4236/me.2017.811085]
23. Sasaki T, Izawa M, Okada Y. Current trends in health insurance systems: OECD countries vs Japan. Neurol Med Chir (Tokyo) 2015;55(4):267-275 [FREE Full text] [doi: 10.2176/nmc.ra.2014-0317] [Medline: 25797778]
24. González A, Darby S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 2004 Jan 31;363(9406):345-351. [doi: 10.1016/S0140-6736(04)15433-0] [Medline: 15070562]
25. Japan Health Policy Now. Payment System URL: http://japanhpn.org/en/finan2/ [accessed 2019-10-06]
26. Japan Council for Quality Health Care. 2018. URL: https://jcqhc.or.jp/wp-content/uploads/2018/10/20181005.pdf [accessed 2019-10-06]
27. Yagi N, Mackey TK, Liang BA, Gerlt L. Policy review: Japan-Philippines Economic Partnership Agreement (JPEPA)--analysis of a failed nurse migration policy. Int J Nurs Stud 2014;51(2):243-250. [doi: 10.1016/j.ijnurstu.2013.05.006] [Medline: 23787219]
28. Efendi F, Mackey TK, Huang M, Chen C. IJEPA: Gray area for health policy and international nurse migration. Nurs Ethics 2017 May;24(3):313-328. [doi: 10.1177/0969733015602052] [Medline: 26385904]
29. The Japan Times. 2018 Mar 28. Number of Foreign Residents in Japan Hit Record 2.56 Million at End of 2017 URL: https://www.japantimes.co.jp/news/2018/03/28/national/number-foreign-residents-japan-hit-record-high-2-56-million-end-2017/#.XRqQbi2ZPUI [accessed 2019-10-06]
30. IQVIA: A New Path to Your Success Via Human Data Science. Top 10 Pharmaceutical Markets Worldwide, 2017 URL: https://www.iqvia.com/-/media/iqvia/pdfs/canada-location-site/top10worldwidesales_en_17.pdf?la=en&hash=C1D59CB0CB2B060CF458F14CF5390D96F3B28238 [accessed 2019-10-06]
31. International Trade Administration. 2016. 2016 Top Markets Report Medical Devices Country Case Study URL: https://www.trade.gov/topmarkets/pdf/Medical_Devices_Japan.pdf [accessed 2019-10-06]
32. Leeming G, Ainsworth J, Clifton DA. Blockchain in health care: hype, trust, and digital health. Lancet 2019 Jun 22;393(10190):2476-2477. [doi: 10.1016/S0140-6736(19)30948-1] [Medline: 31232356]
33. Kuo T, Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 1;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
34. Dilawar N, Rizwan M, Ahmad F, Akram S. Blockchain: securing Internet of Medical Things (IoMT). Int J Adv Comput Sci Appl 2019;10(1). [doi: 10.14569/IJACSA.2019.0100110]
35. Khezr S, Moniruzzaman M, Yassine A, Benlamri R. Blockchain technology in healthcare: a comprehensive review and directions for future research. Appl Sci 2019;9(9):1736 [FREE Full text] [doi: 10.3390/app9091736]

36. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
37. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]
38. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
39. Yogita K. Coindesk. 2019 Feb 13. Pharma Giant Tests IBM Blockchain in Bid to Improve Clinical Trials URL: https://www.coindesk.com/pharma-giant-trials-ibm-blockchain-in-bid-to-improve-clinical-trials [accessed 2019-10-07]
40. AP NEWS. 2019 Apr 22. Susmed: Demonstration of Clinical Data Managing System Using Blockchain Technology URL: https://www.apnews.com/b7aaf7c4c2704b4d91d6497dfe5091fe [accessed 2019-10-07]
41. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
42. Kate H. PR Newswire. 2019 Feb 20. Takeda, Microsoft and EURORDIS Release Report With Action Plan to Help Rare Disease Community Shorten the Diagnostic Odyssey URL: https://www.prnewswire.com/in/news-releases/takeda-microsoft-and-eurordis-release-report-with-action-plan-to-help-rare-disease-community-shorten-the-diagnostic-odyssey-832176691.html [accessed 2019-10-07]
43. McCurry J. Former Novartis employee arrested over valsartan data. Lancet 2014 Jun 21;383(9935):2111. [doi: 10.1016/s0140-6736(14)61015-1] [Medline: 24959638]
44. Clauson K, Breeden EA, Davidson C, Mackey TK. Leveraging blockchain technology to enhance supply chain management in healthcare. Blockchain in Healthcare Today 2018 Mar 23;1 [FREE Full text] [doi: 10.30953/bhty.v1.20]
45. International Institute of Research Against Counterfeit Medicines. 2017 Feb 10. JAPAN: More counterfeit hepatitis C Medicines Discovered URL: https://www.iracm.com/en/2017/02/japan-counterfeit-hepatitis-c-medicines-discovered/ [accessed 2019-10-07]
46. Sato D. [Counterfeit medicines--Japan and the world]. Yakugaku Zasshi 2014;134(2):213-222 [FREE Full text] [doi: 10.1248/yakushi.13-00230-3] [Medline: 24492225]
47. Khan MH, Tanimoto T, Nakanishi Y, Yoshida N, Tsuboi H, Kimura K. Public health concerns for anti-obesity medicines imported for personal use through the internet: a cross-sectional study. BMJ Open 2012;2(3):pii: e000854 [FREE Full text] [doi: 10.1136/bmjopen-2012-000854] [Medline: 22581794]
48. Japan Pharmaceutical Manufacturers Association. 2014. Counter Measures Against Counterfeit Medicines URL: http://www.jpma.or.jp/english/globalhealth/statement/fake.html [accessed 2019-10-07]
49. Olga K. The Japan Times. 2018 Apr 27. IBM Ties Up With Asahi Refining and Others on Blockchain Tech to Track Jewels From Mine to Store URL: https://www.japantimes.co.jp/news/2018/04/27/business/tech/ibm-ties-asahi-refining-others-blockchain-tech-track-jewels-mine-store/#.XR0zGS2ZOXE [accessed 2019-10-07]
50. Cabinet Office. Health and Medical Field URL: https://www8.cao.go.jp/kisei-kaikaku/english/pdf/140613/item2-2.pdf [accessed 2019-10-07]
51. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: systematic review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
52. Matsumoto T, Honda M. The evaluation of the health information exchange with the number of usage and the introduction of outpatient to the hospitals at Nagasaki Japan. Stud Health Technol Inform 2017;245:1373. [Medline: 29295452]
53. Halamka J. CICS. 2011. Addressing Japan's Healthcare Challenges With Information Technology URL: https://csis-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/publication/110830_Halamka_AddressingJapanHealthcare_Web.pdf [accessed 2019-10-07]
54. Abraham C, Nishihara E, Akiyama M. Transforming healthcare with information technology in Japan: a review of policy, people, and progress. Int J Med Inform 2011 Mar;80(3):157-170. [doi: 10.1016/j.ijmedinf.2011.01.002] [Medline: 21292546]
55. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
56. Taipei Times. 2018 Sep 1. Hospital Launches Online Platform URL: http://www.taipeitimes.com/News/taiwan/archives/2018/09/01/2003699590 [accessed 2019-10-07]
57. Nicky M. Ledger Insights. 2018. Taiwanese Hospital Launches Blockchain Health Records URL: https://www.ledgerinsights.com/blockchain-health-records-taiwan/ [accessed 2019-10-07]
58. Lin L, Warren-Gash C, Smeeth L, Chen P. Data resource profile: the National Health Insurance Research Database (NHIRD). Epidemiol Health 2018;40:e2018062 [FREE Full text] [doi: 10.4178/epih.e2018062] [Medline: 30727703]
59. e-estonia. Healthcare URL: https://e-estonia.com/solutions/healthcare/e-health-record/ [accessed 2019-10-07]
60. Taavi E. Nortal. 2018 Feb 21. Blockchain and healthcare: the Estonian experience URL: https://nortal.com/blog/blockchain-healthcare-estonia/ [accessed 2019-10-07]

61. MIT Digital Currency Initiative. 2016. MedRec: A Case Study for Blockchain in Healthcare URL: https://dci.mit.edu/research/blockchain-medical-records [accessed 2019-10-07]
62. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
63. Pirtle C, Ehrenfeld J. Blockchain for healthcare: the next generation of medical records? J Med Syst 2018 Aug 10;42(9):172. [doi: 10.1007/s10916-018-1025-3] [Medline: 30097733]
64. Yano K. The medical doctor qualification certificate project by the Japan medical association. Japan Med Assoc J 2014 Apr 1;57(2):65-67 [FREE Full text] [Medline: 25624780]
65. HubSpot. 2018. Improving Provider Data Accuracy URL: https://cdn2.hubspot.net/hubfs/4801399/18-SYN-001-Synaptic-Website/downloads/Synaptic_Health_Alliance_Blockchain_White_Paper.pdf [accessed 2019-10-07]
66. Sweeney E. Fierce Healthcare. 2018. Aetna, Ascension Join Blockchain Alliance Targeting Provider Directories URL: https://www.fiercehealthcare.com/tech/aetna-ascension-join-blockchain-alliance-targeting-provider-directories [accessed 2019-10-07]
67. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
68. Gordon WJ, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
69. Kyodo. The Japan Times. 2015 Oct 7. Woman in Her 70s First Confirmed Victim of My Number-related Fraud URL: https://www.japantimes.co.jp/news/2015/10/07/national/crime-legal/kanto-woman-70s-first-confirmed-victim-fraud-laid-number-system/#.XR0rnS2ZOXE [accessed 2019-10-07]
70. Silver L. Pew Research Center. 2019 Feb 5. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-10-07]
71. Dwivedi A, Srivastava G, Dhar S, Singh R. A decentralized privacy-preserving healthcare blockchain for IoT. Sensors (Basel) 2019 Jan 15;19(2):pii: E326 [FREE Full text] [doi: 10.3390/s19020326] [Medline: 30650612]
72. Panarello A, Tapas N, Merlino G, Longo F, Puliafito A. Blockchain and IoT integration: a systematic survey. Sensors (Basel) 2018 Aug 6;18(8) [FREE Full text] [doi: 10.3390/s18082575] [Medline: 30082633]
73. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where is current research on blockchain technology?-a systematic review. PLoS One 2016;11(10):e0163477 [FREE Full text] [doi: 10.1371/journal.pone.0163477] [Medline: 27695049]
74. Ito J, Edirippulige S, Aono T, Armfield N. The use of telemedicine for delivering healthcare in Japan: systematic review of literature published in Japanese and English languages. J Telemed Telecare 2017 Oct 28;23(10):828-834 [FREE Full text] [doi: 10.1177/1357633x17732801]
75. Armstrong S. Bitcoin technology could take a bite out of NHS data problem. Br Med J 2018:k1996 [FREE Full text] [doi: 10.1136/bmj.k1996]
76. Kobayashi M. The Japan Times. 2018 Sep 18. Telemedicine via Smartphone Apps Gaining in Popularity in Japan URL: https://www.japantimes.co.jp/news/2018/09/18/national/science-health/telemedicine-via-smartphone-apps-gaining-popularity-japan/#.XR-FyXt7mXG [accessed 2019-10-07]
77. Siripala T. The Japan Times. 2018 Jun 9. Japan's Robot Revolution in Senior Care URL: https://www.japantimes.co.jp/opinion/2018/06/09/commentary/japan-commentary/japans-robot-revolution-senior-care/#.XR76ky2ZN24 [accessed 2019-10-07]
78. The Economist. 2017 Nov 23. Japan is Embracing Nursing-Care Robots URL: https://www.economist.com/business/2017/11/23/japan-is-embracing-nursing-care-robots [accessed 2019-10-06]
79. Reuters. 2018 Mar 27. Aging Japan: Robots May Have Role in Future of Elder Care URL: https://www.reuters.com/article/us-japan-ageing-robots-widerimage/aging-japan-robots-may-have-role-in-future-of-elder-care-idUSKBN1H33AB [accessed 2019-10-06]
80. Volker S, Eduardo CF, Marco D. Managing Byzantine Robots via Blockchain Technology in a Swarm Robotics Collective Decision Making Scenario. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. 2018 Presented at: AAMAS'18; July 10-15, 2018; Stockholm, Sweden p. 541-549 URL: https://pdfs.semanticscholar.org/e5f2/e9d3b89c1496bebf42da26e73fb6a767ac12.pdf
81. Ferrer EC. The Blockchain: A New Framework for Robotic Swarm Systems. In: Proceedings of the Future Technologies Conference. 2018 Presented at: FTC'18; Novermber 15-16, 2018; Vancouver p. 1037-1058 URL: https://arxiv.org/pdf/1608.00695v1.pdf
82. Forbes. 2019 Jun 26. Japan’s Blockchain Sandbox Is Paving The Way For The Fintech Future URL: https://www.forbes.com/sites/japan/2019/06/26/japans-blockchain-sandbox-is-paving-the-way-for-the-fintech-future/#399dd2233279 [accessed 2019-10-06]
83. Japan External Trade Organization. New Regulatory Sandbox framework in Japan URL: https://www.jetro.go.jp/ext_images/en/invest/incentive_programs/pdf/Detailed_overview.pdf [accessed 2019-10-07]
84. Cabinet Office. 2015. What is the Cross-Ministerial Strategic Innovation Promotion Program? URL: https://www8.cao.go.jp/cstp/panhu/sip_english/5-8.pdf [accessed 2019-10-06]

85. Nasdaq. 2016 Oct 31. The Blockchain Developer Shortage: Emerging Trends and Perspectives URL: https://www.nasdaq.com/article/the-blockchain-developer-shortage-emerging-trends-and-perspectives-cm701294 [accessed 2019-10-06]
86. Fintechnews Singapore. 2018 Oct 30. Japan's Banks Elect Fujitsu to Build Their Blockchain for Interbank Settlements URL: http://fintechnews.sg/25322/blockchain/fujitsu-blockchain-japan-bank-interbank-transactions/ [accessed 2019-10-06]
87. Yong J. CCN: News and Opinions. 2018 Jan 17. Japan's Largest Bank MUFG Will Open Cryptocurrency Exchange: Local Reports URL: https://www.ccn.com/japans-largest-bank-mufg-will-open-cryptocurrency-exchange-local-media-reports/ [accessed 2019-10-06]
88. The Straits Times. 2016 Nov 15. OCBC First in S-E Asia to Use Secure Blockchain Technology URL: https://www.straitstimes.com/business/banking/ocbc-first-in-s-e-asia-to-use-secure-blockchain-technology [accessed 2019-10-06]
89. MIT Digital Currency Initiative. 2018. Cellular Structure for a Digital Fiat Currency URL: https://dci.mit.edu/research/2018/7/26/cellular-structure-for-a-digital-fiat-currency [accessed 2019-10-07]
90. CoinTelegraph. 2018 Jul 18. Japan Revamps Financial Regulatory Agency to Address Issues in FinTech, Cryptocurrencies URL: https://cointelegraph.com/news/japan-revamps-financial-regulatory-agency-to-address-issues-in-fintech-cryptocurrencies [accessed 2019-10-06]
91. e-estonia. e-estonia Blockchain URL: https://e-estonia.com/category/blockchain/ [accessed 2019-10-05]

Abbreviations

eHealth: electronic health
EHR: electronic health record
GDP: gross domestic product
HIE: health information exchange
IEEE: Institute of Electrical and Electronics Engineers
IoT: Internet of Things
IoMT: Internet of Medical Things
JCQHC: Japan Council for Quality Health Care
mHealth: mobile health
MHLW: Ministry of Health, Labour and Welfare
OECD: Organisation for Economic Co-operation and Development
SIP: Cross-ministerial Strategic Innovation Promotion Program
IT: information technology
UHC: universal health coverage

Edited by P Zhang, K Clauson; submitted 06.02.19; peer-reviewed by R Hylock, M Káplár, A Fittler; comments to author 27.04.19; revised version received 17.07.19; accepted 17.09.19; published 09.01.20.

Please cite as:
Mackey T, Bekki H, Matsuzaki T, Mizushima H
Examining the Potential of Blockchain Technology to Meet the Needs of 21st-Century Japanese Health Care: Viewpoint on Use Cases and Policy
J Med Internet Res 2020;22(1):e13649
URL: https://www.jmir.org/2020/1/e13649
doi:10.2196/13649
PMID:31917371

Original Paper

Visualizing an Ethics Framework: A Method to Create Interactive Knowledge Visualizations From Health Policy Documents

Joanna Sleigh[1*], MA; Manuel Schneider[1*], MSc; Julia Amann[1], PhD; Effy Vayena[1], PhD

Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Joanna Sleigh, MA
Health Ethics and Policy Lab
Department of Health Sciences and Technology
ETH Zurich
Hottingerstrasse 10
Zurich
Switzerland
Phone: 41 44 633 99 31
Email: joanna.sleigh@hest.ethz.ch

Abstract

Background: Data have become an essential factor in driving health research and are key to the development of personalized and precision medicine. Primary and secondary use of personal data holds significant potential for research; however, it also introduces a new set of challenges around consent processes, privacy, and data sharing. Research institutions have issued ethical guidelines to address challenges and ensure responsible data processing and data sharing. However, ethical guidelines directed at researchers and medical professionals are often complex; require readers who are familiar with specific terminology; and can be hard to understand for people without sufficient background knowledge in legislation, research, and data processing practices.

Objective: This study aimed to visually represent an ethics framework to make its content more accessible to its stakeholders. More generally, we wanted to explore the potential of visualizing policy documents to combat and prevent research misconduct by improving the capacity of actors in health research to handle data responsibly.

Methods: We used a mixed methods approach based on knowledge visualization with 3 sequential steps: qualitative content analysis (open and axial coding, among others); visualizing the knowledge structure, which resulted from the previous step; and adding interactive functionality to access information using rapid prototyping.

Results: Through our iterative methodology, we developed a tool that allows users to explore an ethics framework for data sharing through an interactive visualization. Our results represent an approach that can make policy documents easier to understand and, therefore, more applicable in practice.

Conclusions: Meaningful communication and understanding each other remain a challenge in various areas of health care and medicine. We contribute to advancing communication practices through the introduction of knowledge visualization to bioethics to offer a novel way to tackle this relevant issue.

(J Med Internet Res 2020;22(1):e16249) doi:10.2196/16249

KEYWORDS

ethics framework; health data; health policy; knowledge visualization; systems map

Introduction

We live in an era where data are omnipresent and seemingly omnipotent. Data constitute one of the forces, if not *the* driving force, behind personalized and precision medicine. Traditional health data sources such as medical records and clinical trial data are nowadays complemented by an ever-increasing amount of behavioral and lifestyle data, which we create by interacting with everyday technologies such as our smartphones. Although data collection remains important, data sharing is fundamental to modern scientific practice and is of great value to the health sciences. First, this is because data are crucial to the confirmation of research findings and the replication of results [1]. Second, making data available enables scientists to collaborate and build on the work of others [2]. Third, reusing

data enables researchers to leverage research investments, particularly public funding [3]. Fourth, data sharing is integral to the advancement of research and innovation [4]. In the health sciences specifically, data sharing has the potential to transform health care and inform clinical research; quality measurement; and, ultimately, public safety [5].

However, although the quantity and types of data available for research are rapidly expanding, the handling of such data is a complex process that involves and impacts several stakeholder groups such as patients and research institutions. In recognition of precision medicine's reliance on big data, the Swiss Personalized Health Network (SPHN) produced the Ethical Framework for the Responsible Usage of Personal Data in Health Research [6]. The framework guides SPHN's actors (such as researchers) as they endeavor to handle data ethically, to inform research participants about these ethical practices, and to tackle concerns regarding privacy and misuse of data. However, the SPHN ethical framework is innately complex. Like many health policy documents, the SPHN framework describes a multilayered, nonlinear process that involves several stakeholders. The problem with this is that the document needs to refer to other elements of the framework to cover 1 aspect fully [7]. For example, the process of consent involves the research participant, the researcher, and the institution. In addition, it is a process interwoven at different stages of a research project, from the very beginning to long after the project has finished.

Ethical and policy guidance is only as effective as its application. To combat and prevent research misconduct and to foster data sharing practices, stakeholders involved need to understand available guidance and apply it in practice. Some researchers have questioned the efficacy of policy documents as communication tools [8]. Research shows that senior decision makers often do not read long policy documents [9]. However, senior decision makers are not the only critical audience for health policy. For example, the SPHN ethical framework is relevant to diverse members of the SPHN network, from researchers to medical practitioners and individuals who participate in studies. Policy documents are no match to the challenge of communicating complex information to tremendously diverse audiences [10], and well-intentioned stakeholders can find it onerous to act on the information in the document. Whenever this happens, policy documents defy their purpose.

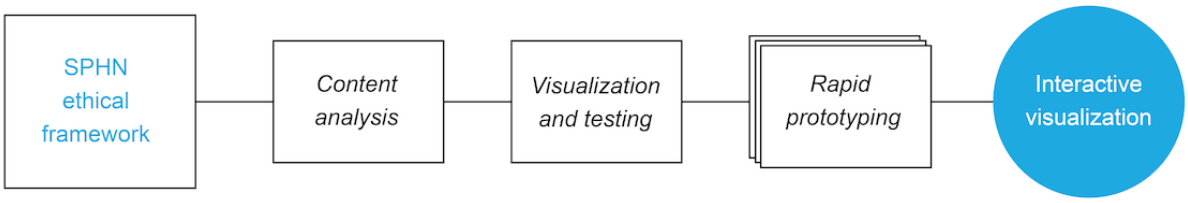
So how can we ensure that policy documents actually fulfill their purpose? Extensive research exists on how to communicate complex information to stakeholders. One promising approach is knowledge visualization. Knowledge visualization "examines the use of complementary visual representations to improve the transfer and creation of knowledge between at least two persons" [11]. Evidence indicates that the active integration of visual representations improves learning significantly [12,13]. Theories that support this include Paivio's dual coding theory [14], which asserts that we process verbal information and pictorial information in different cognitive systems, as well as Chandler and Sweller's cognitive load theory [15], which argues that multiple sources of information facilitate learning by reducing working memory. These insights and opportunities are currently not used to inform knowledge dissemination in health policy making [10,16,17]. To address this gap, we adopted a knowledge visualization methodology and applied a mixed methods approach to translating an existing ethical framework into an interactive knowledge visualization tool. In this study, we present our methodological approach and describe the process that led us to this method. To do this, we take the SPHN ethical framework for the Responsible Usage of Personal Data in Health Research as a case study.

Methods

Overview

Our approach to transforming a complex framework into an interactive visualization involved 3 steps, as shown in Figure 1. To begin with, we conducted a quantitative content analysis to distill and structure the knowledge inside the document. We used a combination of inductive and deductive methods to create a conceptual representation of the content sequentially. In a second step, we transferred the conceptual data into 4 different visual forms and, then, tested these visualizations through expert review to select the most promising candidate. Although the visualizations did already make the content of the framework more accessible, the interwoven structure of the knowledge was still not entirely represented. In a final step, we, therefore, iteratively developed an interactive version of the visualization through rapid prototyping. We look now at each of these steps in more detail.

Figure 1. This figure gives an overview of the method and outlines the 3 primary steps: content analysis, visualization, and rapid prototyping. The rectangle on the left represents the Swiss Personalized Health Network's ethical framework in its original form as a document, and the circle on the right represents the interactive visualization of the framework's content. SPHN: Swiss Personalized Health Network.



Content Analysis

Our first step was to identify the knowledge and knowledge structure manifested in the framework. To do this, we analyzed the framework using a qualitative content analysis approach that combined inductive and deductive category development [18-20]. Moreover, 1 author (JS) with training and previous experience in inductive and deductive coding performed the analysis. We ensured intracoder reliability through 3 rounds of coding, with the first 2 within a month and the third round 3 months later. We resolved any discrepancies in discussions within the research team.

We conducted an inductive analysis to identify (1) the key elements in the ethical framework, (2) stakeholders, (3) the knowledge types present in the SPHN framework, and (4) the connections between the elements and stakeholders. We started with open coding, with the restriction that codes had to be mutually exclusive to ensure that the resulting knowledge structure remains unambiguous [21].

We then grouped codes that concerned the same subject matter and merged them into categories. To give an example, we merged the codes *withdraw* and *revoking consent* into *withdrawal process* just as we merged *communication* and *information* into *participant information*. These categories form the lowest layer of abstraction in the knowledge structure of the framework, that is, the subthemes. This task was conducted iteratively, where we revised, refined, and checked the subthemes to ensure that they remained mutually exclusive.

In the next step, we formed overarching themes out of the subthemes, adding another layer of abstraction. As an example, the 2 subthemes *further use* and *withdrawal process* both belong to the theme *consent process*. These groups can overlap as a subtheme can belong to multiple groups. For example, *further use* is part of both the *consent process* and the *data and samples* themes.

We then determined the primary stakeholders from the frequency of their occurrences in the text and used axial coding to identify and map out the relationships between the stakeholders, themes, and subthemes [22]. Although some of the coded elements explicitly indicated relationships among each other and with stakeholders, we had to infer others from context.

Finally, we used deductive analysis to assign the stakeholders, themes, subthemes, and their relationships with 1 or more of the 4 normative ethical principles of the SPHN framework: respect for persons, data fairness, privacy, and accountability. The resulting 4 groups are not mutually exclusive as a stakeholder, theme, and subtheme can be affected by several ethical principles. This fact contributed to the complexity of the original SPHN framework document.

A total of 3 content experts from the fields of bioethics (EV and AB) and public health (FG) reviewed the coding and the knowledge structure. We selected the experts through purposive sampling. Moreover, 2 of the 3 experts (EV and AB) contributed to the development of the original SPHN framework and made sure that the result reflects the entirety of the framework's content.

Multimedia Appendices 1-4 present the results of the inductive and deductive analysis.

Visualization Methods

To visualize the previously derived knowledge structure in a simple yet comprehensive way, we tested different graphics and visualization methods: alluvial diagrams, graphics such as symbols, concept maps, and systems maps. The 4 methods were chosen based on Burkhard's model of visualization types for knowledge visualization [23]. The following paragraphs describe the visualization methods. See Multimedia Appendix 5 for the outputs.

Alluvial Diagram

Alluvial diagram is a type of flow diagram, or branch-based diagram, that represents weighted correlations between categorical dimensions, visually linking the number of elements to shared categories [24]. As abstract and schematic representations, alluvial diagrams are used to explore structural relationships among parts and are, therefore, used to explain concepts and reduce complexity. For this reason, we first employed an alluvial diagram to visually explore the various relationships among themes, actors, and ethical principles.

Signs, Symbols, and Sketches

Sketches, drawings, symbols, and icons are nonverbal representational forms used for knowledge transfer and communication. The making of meaning from visual representations is a very different undertaking than that of language [25]. Sketches are, thus, a useful and powerful visualization tool that enables quick communication and that stimulates creativity by leaving room for interpretation [26]. For these reasons, we transmediated the key themes and subthemes into a series of icons and visual metaphors to retain user attention, enhance understanding, and improve recall.

Concept Map

Concept mapping, also referred to as structured conceptualization, is an established method for the organization and representation of knowledge. The method produces a map that consists of nodes and lines—the nodes indicating concepts and the connecting lines denoting relationships between them [27]. What differentiates concept mapping from methods such as mind mapping, cluster mapping, and flow charts is that concept mapping uses a top-down structure to show the relationships between themes and subthemes with the overall concept. In this study, we chose concept mapping to visualize the ethical framework's 4 core ethical principles and the actors, issues, and concepts that relate to them.

Systems Map

Systems mapping is the process of visually representing and describing an entire system, including the elements and actors involved as well as their relationships, links, and interconnections. This method makes clear how things such as information or materials flow through a system. Other types of systems maps are causal loop diagrams, actor-network maps, and value chain maps. Reasons for using this method are as follows: (1) systems maps make sense of complexity, (2) they engage stakeholders by highlighting their position in the system,

and (3) they enable both issues and opportunities to be easily identified [28]. We chose this method to visualize the entirety of the ethical framework's knowledge system to make clear the patterns of process and underlying relationships between values and beliefs (mental models) of the actors (people and organizations) involved, responsible and impacted.

Testing Through Expert Review

We invited 2 experts (EV and AB) to assess which visualization method was most appropriate for the SPHN framework content. Both experts were selected because they had contributed to the revision of the original SPHN framework in 2018 and the underlying analysis of existing policies [29]. We conducted the expert reviews as informal interviews, wherein we presented the 2 reviewers each of the tested visualizations to determine their respective strengths and weaknesses. To do so, we structured interviews based on Burkhard's knowledge visualization framework [23]:

1. Attention: Is the visualization attractive and engaging?
2. Context: Does the visualization convey why the knowledge is needed and is of value?
3. Overview: Does the visualization give an overview of the complexity of the framework?
4. Options to act: Does the visualization provide options to act, to use, and to apply the knowledge.
5. Details: Is the amount of detail appropriate?

The experts' review showed that although many of the visualization methods were engaging, they oversimplified the content, making important details inaccessible. For example, the concept maps were determined to be effective at showing the actors and elements inherent to each ethical principle, but neither did they explain how the principles related to each other nor did they provide definitions. Only the system map was successful in providing both an overview and adequate detail of the ethical framework. However, according to the experts, the systems map failed to make the underlying ethical principles visible. Despite some shortcomings, we identified the systems map to be the most appropriate method.

Rapid Prototyping

To address the shortcomings of the systems map and make the SPHN ethical framework content accessible in its entirety, we developed an interactive visualization using rapid prototyping [30]. We wanted to enable users to explore across registers, from the big concepts to the specific details in the ethical framework's range of knowledge types (declarative, experiential, individual, orientational, and procedural). This goal was informed by Ausubel's assimilation theory [31]. We, hence, devised a Web-based system map that proceeded from the more general, more inclusive concepts to the more specific information.

To assess the design and usability of the prototype, we conducted 3 expert reviews (by DG, IS, and ML) at different stages of the prototyping process with a different reviewer for each stage. We used convenience sampling to find experts for user experience (UX) research, design, and storytelling. For this purpose, we presented reviewers with the prototype and asked them to provide feedback. A UX researcher, with a background in media studies and economics, participated in the first expert review. A UX designer with a Masters in Design and over 15 years of experience in product development undertook the second review. This expert was familiar with the topic of data processing, yet had no experience in the health sector. A UX copywriter completed the final review. This reviewer has a background in nanotechnology design and is also familiar with data processing and data sharing practices. These experts judged the prototypes according to usability requirements and according to the previously introduced questions by Burkhard [23]. We systematically recorded and compared the comments and suggestions for improvement received from each of the 3 experts to inform the next prototype iteration.

To prototype an interactive systems map, we used a data visualization platform called Kumu [32]. We first loaded the identified themes, subthemes, and stakeholders into Kumu and designated the nodes by labels and size. Large nodes represented stakeholders, medium-sized nodes indicated themes, and we gave small nodes to subthemes. We then clustered and linked these nodes according to theme and subtheme hierarchies and the relationships identified from axial coding. Afterward, we pollinated the nodes with metadata and descriptions derived directly from the SPHN ethical framework. We simplified the wording at times or added definitions from the SPHN ethical framework glossary document. Concurrently, we embedded the visual icons developed during the initial testing phase to support content recall for users. For the primary 11 themes (consent, upholding human rights, authorization procedures, governance structures, data + sample processing, accountability processes, security control processes, scientific research, data and samples, transparency, and sharing process), we animated these line icons into gifs that loop.

Expert reviewers noted some limitations with Kumu; thus, to overcome these, we developed a website with a customized interactive visualization of the nodes with additional functionality and information. In addition, we implemented custom views that present content based on the different stakeholders' perspectives on the systems map. Each perspective was composed of 11 views that highlight the themes and explain their importance according to the respective stakeholder's responsibilities and interests. Furthermore, we integrated the 4 ethical principles by allowing the user to highlight the affiliation of the nodes to the activated principle.

Results

In this section, we present the final interactive visualization prototype of the SPHN ethical framework and its features. In total, we identified 4 primary stakeholders, 3 general themes, 8 research process themes, 2 effect themes, and 30 subthemes. The inductive analysis further revealed that the SPHN ethical framework comprises multiple knowledge types. These included the following: declarative knowledge (know-about), experiential knowledge (know-why, eg, causes), individual knowledge (know-who), orientational knowledge (know-where), and procedural knowledge (know-how) [33]. Multimedia Appendices 1-5 present the results of intermediate steps in more detail.

The focal point of the prototype is the interactive systems map visualizing the SPHN ethical framework. It shows the stakeholders, themes, subthemes, and their relations to each other. The map is made up of edges and nodes. The large-sized nodes correspond to the stakeholders, whereas the medium-sized nodes represent the themes and the small-sized nodes indicate the subthemes. In addition to the map itself, there are complementary functionalities, which we describe along with the other interaction techniques in the following paragraphs.

First, users can hover to highlight a node and its connections (see Figure 2). This functionality is integral for highlighting the interconnectedness of the themes and stakeholder.

Second, the select function is used to activate the display of a node's information and metadata. In other words, by clicking on a node, users can access the information based on the SPHN ethical framework (see Figure 2). Each node is made up of a title, a definition or description of what it refers to, information about why it matters, and references to the connected nodes. In addition, to make the information easier to understand, each node includes embedded explanatory media such as animated gifs, videos, and graphics. Figure 3 exhibits a screenshot of the final interactive visualization [34] (additional screenshots are provided in Multimedia Appendix 6).

Third, to enable navigation of the content according to the stakeholder categorizations, stakeholder perspectives were defined and can be activated through buttons above the map. Each stakeholder's perspective is divided into 11 subperspectives, titled parts, reflecting the 11 themes of the respective perspective. By clicking on a perspective and subperspective button, for example, *Researcher* and *Part 1*, the nodes corresponding to the subperspective are highlighted in the map, and additional information for the selection is displayed (see Figure 2). These perspectives enable a linear navigation approach to the clusters and connections of the systems map. We further adjusted the language style and level of detail according to the stakeholder's respective responsibilities and interests. For example, the society viewpoint for *data governance* is less detailed than the institution viewpoint for the same topic.

Fourth, to incorporate the SPHN ethical framework's 4 ethical principles, we used a color legend. As such, we assigned each ethical principle a color and tagged the nodes accordingly. By clicking an ethical principle in the legend, users can then highlight the relevant nodes (see Figure 2). In this way, users experience the interwoven nature of ethical principles.

Figure 2. In image 2.1, we see the effect of hovering over a node that highlights connections. Image 2.2 shows the function of clicking to access information and metadata. Image 2.3 depicts the buttons that represent the stakeholder perspectives and the node highlighting in the map below as response when one of them is activated. Image 2.4 demonstrates the function of highlighting ethical principles using the color legend simultaneously to the other functions where each color corresponds to an ethical principle.

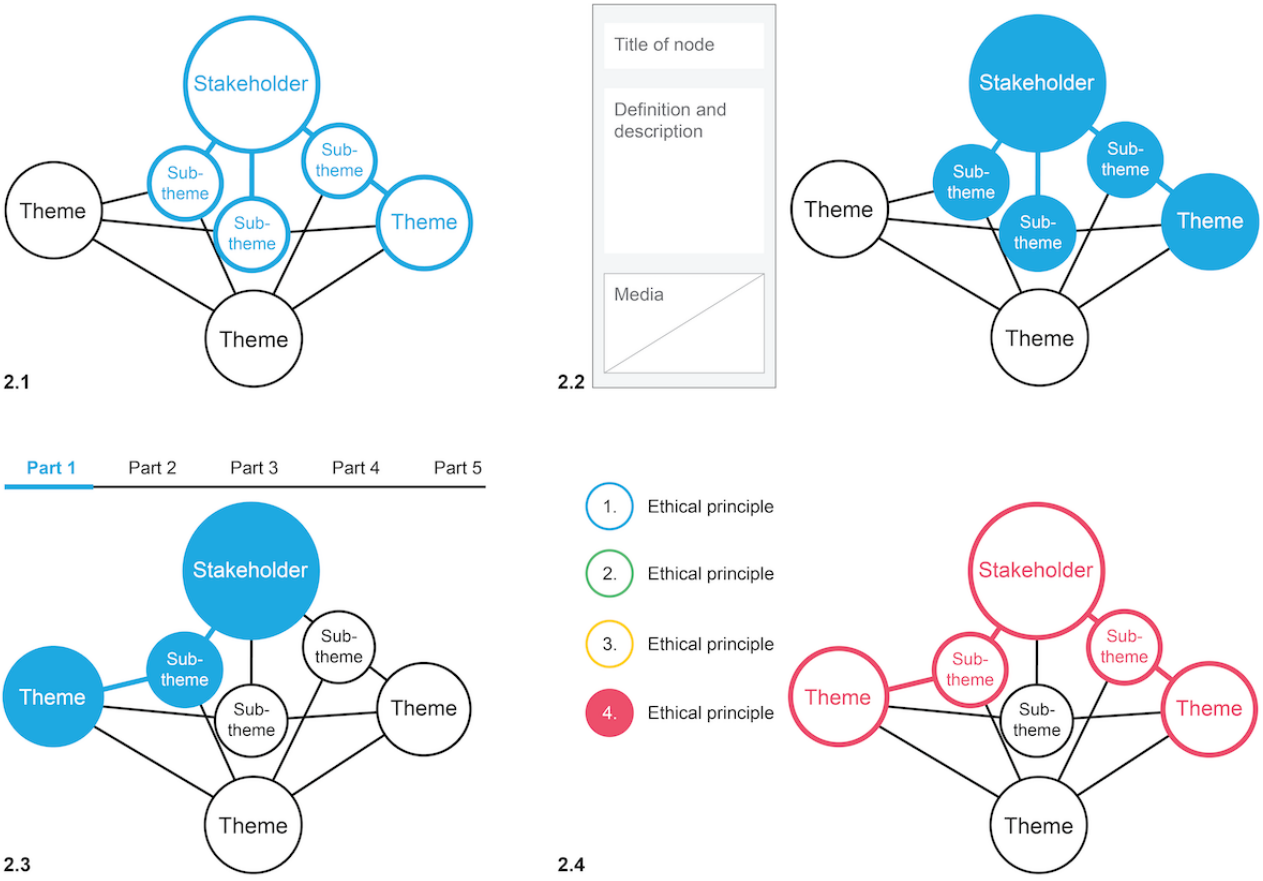
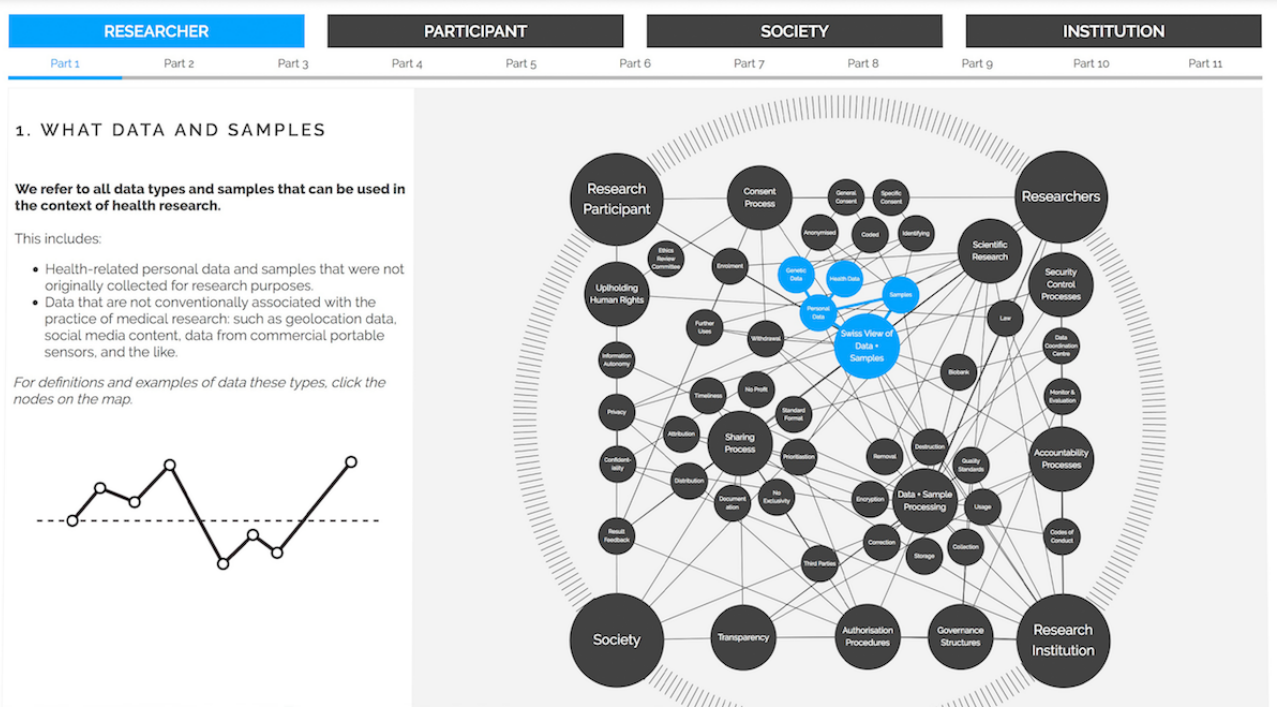


Figure 3. Screenshot of the interactive systems map.



Discussion

Principal Findings

In this study, we developed an interactive visualization to navigate the content of the SPHN ethical framework. Our mixed methods approach consisted of qualitative analysis, visualization techniques, and rapid prototyping. In contrast to the ethical framework's original form as a text document, our interactive visualization offers users an overview of the ethical framework content and provides access to detailed descriptions and definitions enriched with multimedia content. Another unique function of our interactive visualization is that it allows users to examine the relationships between elements and themes.

Our method finds application beyond the specific type of policy document we used in our case study. Specifically, the individual parts of the method can be applied independent of context to transform a defined text scope. Only the choice of visualization type based on the expert review was specific to bioethics. This dependency can be resolved by calling in experts from other respective fields. The method presented in this study, therefore, holds great potential for a variety of text contents beyond policy documents and bioethics.

Limitations

Future research is needed to unlock the potential of this visualization approach. To begin with, this study does not assess the interactive visualization's educational powers compared with the original policy document. Further research is, thus, required to measure the effectiveness of the knowledge transfer. In addition, systematic user testing is needed to resolve functional shortcomings. The combination of these 2 assessments would optimize the tool's efficacy and, therefore, improve the impact of data sharing policies on the stakeholders' actions.

Another limitation is that the visualization does not assist the stakeholders' understanding of the dependencies between different elements of the framework. In the words of Tufte, understanding "... is to know what cause provokes what effect, by what means, at what rate" [35]. To resolve this, a functionality that highlights cause and effect could be incorporated and tested. For example, a game function that allows the user to remove nodes from the map and then to see how the scenario evolves enables the stakeholders to explore the impact of different elements of the framework in an engaging way.

Conclusions

This study sheds light on how to develop an interactive visualization from a policy document and lays the foundation to innovate ethical framework dissemination practices on the Web. In practice, this method to translate ethical frameworks gives bioethicists and health care policy makers a tool to communicate complex information to diverse audiences. More broadly, the results offer guidance to researchers, practitioners, and designers who create dynamic visualization for scientific and scholarly communication.

Meaningful communication and understanding each other remain a challenge in various areas of health care and medicine. We contribute to advancing communication practices through the introduction of knowledge visualization to bioethics to offer a novel way to tackle this relevant issue. Our work bears value for every person involved in modern health research: from policy makers who give guidance, to researchers who process data, to the patients who use their smartphones throughout the day and generate data.

Acknowledgments

The authors would like to acknowledge the 2 groups of expert reviewers for their valuable insights and feedback both on the content level and on the early prototypes. The authors would also like to thank Stephanie Bishop for her feedback, proofreading, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Identified framing themes in responsible processing of personal data and samples in health research.
[PDF File (Adobe PDF File), 22 KB - jmir_v22i1e16249_app1.pdf]

Multimedia Appendix 2
Identified research process themes in responsible processing of personal data and samples in health research.
[PDF File (Adobe PDF File), 29 KB - jmir_v22i1e16249_app2.pdf]

Multimedia Appendix 3
Identified effect themes in responsible processing of personal data and samples in health research.
[PDF File (Adobe PDF File), 20 KB - jmir_v22i1e16249_app3.pdf]

Multimedia Appendix 4
Identified stakeholders in responsible processing of personal data and samples in health research.
[PDF File (Adobe PDF File), 25 KB - jmir_v22i1e16249_app4.pdf]

Multimedia Appendix 5
Summary of expert reviews of the visualizations.
[PDF File (Adobe PDF File), 101 KB - jmir_v22i1e16249_app5.pdf]

Multimedia Appendix 6
Screenshots of the interactive visualization.
[PDF File (Adobe PDF File), 2822 KB - jmir_v22i1e16249_app6.pdf]

References

1. Munafò MR, Nosek BA, Bishop DV, Button KS, Chambers CD, du Sert NP, et al. A manifesto for reproducible science. Nat Hum Behav 2017;1(21). [doi: 10.1038/s41562-016-0021]
2. Louis KS, Jones LM, Campbell EG. Sharing in science. Am Sci 2002;90(4):304. [doi: 10.1511/2002.4.304]
3. Pilat D, Fukasaku Y. OECD principles and guidelines for access to research data from public funding. Data Sci J 2007;6:OD4-O11. [doi: 10.2481/dsj.6.od4]
4. Borgman CL. The conundrum of sharing research data. Acta Anaesthesiol Scand 2012;63(6):1059-1078. [doi: 10.1002/asi.22634]
5. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. Bull World Health Organ 2018;96(1):66-68. [doi: 10.2471/blt.17.197426]
6. Ethical framework for responsible data processing in personalized health research Bern.: Swiss Personalised Health Network; 2018. URL: https://www.sphn.ch/dam/jcr:6fb78ffa-95c8-4372-bfb1-5c9b1e2cb53d/Ethical_Framework_20180507_SPHN.pdf [accessed 2018-07-01]
7. Buse K, Mays N, Walt G. The health policy framework. In: Making Health Policy. Second Edition. Berkshire: Open University Press; 2012:4-19.
8. Dagenais C, Ridde V. Policy brief as a knowledge transfer tool: to 'make a splash', your policy brief must first be read. Gac Sanit 2018;32(3):203-205 [FREE Full text] [doi: 10.1016/j.gaceta.2018.02.003] [Medline: 29609974]
9. Avey PC, Desch MC. What do policymakers want from us? Results of a survey of current and former senior national security decision makers. Int Stud Q 2014;58(2):227-246. [doi: 10.1111/isqu.12111]
10. Zakkar M, Sedig K. Interactive visualization of public health indicators to support policymaking: an exploratory study. Online J Public Health Inform 2017;9(2):e190 [FREE Full text] [doi: 10.5210/ojphi.v9i2.8000] [Medline: 29026455]
11. Burkhard RA, Meier M. Tube map visualization: evaluation of a novel knowledge visualization application for the transfer of knowledge in long-term projects. J Univers Comput Sci 2005;11(4):473-494. [doi: 10.3217/jucs-011-04-0473]

12. Bodemer D, Ploetzner R, Feuerlein I, Spada H. The active integration of information during learning with dynamic and interactive visualisations. Learn Instr 2004 Jun;14(3):325-341. [doi: 10.1016/j.learninstruc.2004.06.006]
13. Fischer F, Bruhn J, Gräsel C, Mandl H. Fostering collaborative knowledge construction with visualization tools. Learn Instr 2002;12(2):213-232. [doi: 10.1016/s0959-4752(01)00005-6]
14. Paivio A. Dual coding theory: retrospect and current status. Can J Psychol 1991;45(3):255-287. [doi: 10.1037/h0084295]
15. Sweller J, Chandler P. Evidence for cognitive load theory. Cogn Instr 1991;8(4):351-362. [doi: 10.1207/s1532690xci0804_5]
16. Bassi J, Lau F. Measuring value for money: a scoping review on economic evaluation of health information systems. J Am Med Inform Assoc 2013;20(4):792-801 [FREE Full text] [doi: 10.1136/amiajnl-2012-001422] [Medline: 23416247]
17. Quinn E, Huckel-Schneider C, Campbell D, Seale H, Milat AJ. How can knowledge exchange portals assist in knowledge management for evidence-informed decision making in public health? BMC Public Health 2014 May 12;14:443 [FREE Full text] [doi: 10.1186/1471-2458-14-443] [Medline: 24884530]
18. Cole FL. Content analysis: process and application. Clin Nurse Spec 1988;2(1):53-57. [doi: 10.1097/00002800-198800210-00025] [Medline: 3349413]
19. Kondracki NL, Wellman NS, Amundson DR. Content analysis: review of methods and their applications in nutrition education. J Nutr Educ Behav 2002;34(4):224-230. [doi: 10.1016/s1499-4046(06)60097-3]
20. Forman J, Damschroder L. Qualitative content analysis. Adv Bioeth 2007;11:39-62. [doi: 10.1016/s1479-3709(07)11003-7]
21. Cho J, Lee E. Reducing confusion about grounded theory and qualitative content analysis: similarities and differences. Qual Rep 2014;19(32):1-20 [FREE Full text]
22. Charmaz K. Constructing Grounded Theory. Second Edition. Los Angeles: Sage Publications Ltd; 2006.
23. Burkhard RA. ETH Research Collection. 2005. Knowledge Visualization URL: https://www.research-collection.ethz.ch/handle/20.500.11850/148745 [accessed 2019-03-02]
24. Sedig K, Parsons P. Design of visualizations for human-information interaction: a pattern-based framework. Synth Lect Vis 2016;4(1):1-185. [doi: 10.2200/s00685ed1v01y201512vis005]
25. Siegel M. More than words: The generative power of transmediation for learning. Can J Educ 1995;20(4):455. [doi: 10.2307/1495082]
26. Langer SK. Philosophy In A New Key. Cambridge: Harvard University Press; 1942.
27. Novak JD, Canas AJ. The theory underlying concept maps and how to construct and use them. Prax Educ 2010;5(1):9-29. [doi: 10.5212/PraxEduc.v.5i1.009029]
28. Karsh BT, Alper SJ. Work system analysis: The key to understanding health care systems. In: Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology). Rockville (MD): Agency for Healthcare Research and Quality (US); 2005:337-348.
29. Blasimme A, Fadda M, Schneider M, Vayena E. Data sharing for precision medicine: policy lessons and future directions. Health Aff (Millwood) 2018 May;37(5):702-709. [doi: 10.1377/hlthaff.2017.1558] [Medline: 29733719]
30. Tripp SD, Bichelmeyer B. Rapid prototyping: an alternative instructional design strategy. Educ Technol Res Dev 1990;38(1):31-44. [doi: 10.1007/bf02298246]
31. Ausubel DP, Novak JD, Hanesian H. Educational Psychology: A Cognitive View. New York: Holt, Rinehart and Winston; 1968.
32. Kumu. URL: https://kumu.io/ [accessed 2019-11-07]
33. Alavi M, Leidner DE. Review: knowledge management and knowledge management systems: conceptual foundations and research issues. Manag Inf Syst Q 2001;25(1):107-136. [doi: 10.2307/3250961]
34. Navigating Health Data Ethics. URL: https://ethicalsystemsmap.hest.ethz.ch/ [accessed 2019-09-01]
35. Tufte ER. Visual Explanations. Cheshire, CT: Graphics Press; 1997.

Abbreviations

SPHN: Swiss Personalized Health Network
UX: user experience

Edited by A Marusic; submitted 14.09.19; peer-reviewed by I Buljan, M Vidak; comments to author 06.10.19; revised version received 17.10.19; accepted 18.10.19; published 14.01.20.

Please cite as:
Sleigh J, Schneider M, Amann J, Vayena E
Visualizing an Ethics Framework: A Method to Create Interactive Knowledge Visualizations From Health Policy Documents
J Med Internet Res 2020;22(1):e16249
URL: https://www.jmir.org/2020/1/e16249
doi:10.2196/16249
PMID:31934866

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org